The Genetic Basis of *Wolbachia*-Induced Male Killing in *Drosophila*

By

Jessamyn Irene Perlmutter

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biological Sciences

February 29, 2020

Nashville, Tennessee

Approved:

Julián Hillyer, Ph.D. (Chair)

Kathy Gould, Ph.D.

Maulik Patel, Ph.D.

Jared Nordman, Ph.D.

Seth Bordenstein, Ph.D. (Advisor)

Dedication

To the millions of flies sacrificed for the science herein

Acknowledgements

Graduate school is a long and humbling experience, and science is not done alone. I am thankful for the many mentors, mentees, peers, friends, family, and others who have helped me build skills and accomplish all that I have.

To my advisor, Seth Bordenstein, who guided me through critical thinking, experimental design, scientific writing, data analysis, professional skill building, and everything else: thank you. I am a better scientist, communicator, and professional because of your guidance and support. Thanks for having an open door, challenging me, and giving honest feedback. You put together a great group of people in the lab and I am glad to be a part of it.

To my other mentors and advisors, thank you. My committee, Julián Hillyer (chair), Kathy Gould, Maulik Patel, and Jared Nordman provided feedback and encouragement over many years that made a big difference in many experiments and helped me see my work from other perspectives. Sarah Bordenstein, the phage WO expert of the world, patiently answered my many questions, provided invaluable technical expertise and brilliant insights into my project, taught me the basics of making phylogenies and using Geneious, and was always willing to chat about the latest tv shows we were watching. Daniel LePage enthusiastically taught me *Drosophila* skills and showed me all the shortcuts and tricks of the trade.

To my fellow labmates past and present, thank you. Many people in the lab have made grad school a better experience. There are too many to name, but I appreciate the help, feedback, conversation, and everything else from the last several years. In particular, I would like to thank Brittany Leigh and Teddy van Opstal for all of the help and advice over the years and for participating in everything.

To my mentees, big thanks to you. Jane Meyers is reliable, a quick learner, and is always willing to do the work that needed to be done, a rare and valuable trait. I always looked forward to our weird conversations on obscure cultural topics and our many failed experimental ideas. Thanks for all of the work you did, and for the fun. Emily Layton impresses everyone with her dedication and her ability to strive for the best quality work. She worked through thick and thin and accomplished an incredible amount that will set her up for whatever path she wants to take. To all of the others who I have taught and mentored in and out of the lab, thank you for being a critical driving force in science.

To my peers and friends, thank you. There are too many to name, but I am grateful for our mutual commiseration and support over the years. There were many people who were essential to various projects and groups that were instrumental experiences in grad school, from outreach to teaching and advising, to going to brunch. I would like to especially thank my fellow movie enthusiast, Katrina Ngo, for all our trips to the theater over the years.

To the groups and resources at Vanderbilt and beyond, thank you. The SyBBURE Searle undergraduate research program was an experience I am very grateful to have had. I met dozens of amazing undergraduate researchers and advisors in the program, and always looked forward to being a part of it. The student groups, the ASM student chapter and IBA, both had incredible scientists and leaders who were dedicated to bridging the gap between scientists and the public, and I am grateful for the opportunities and the partnership of my peers in all of our initiatives. I am also thankful to the many resources and services at Vanderbilt, including the BRET office and ASPIRE program, IGP, the Center for Teaching, and others that helped me build professional skills and get where I am.

To my family, thank you. To my mom and my brother: we are a ridiculous family sometimes, but I am always grateful for the listening ears, moral support, and shared prank gifts and insults. A prophet is not without honor save in her own family. To my grandma, who is always supportive and excited to see me. To my dad, who is no longer here, but never doubted me and was always proud of me.

Finally, I would like to make a special shoutout to the following (incomplete) list for getting me through grad school: my cat, my goldfish, my banana plant, Netflix, Amazon Prime, various food delivery services, coffee, tea, my down feather pillows, Panera, movie theaters and moviemakers everywhere, my Crock-Pot, Twitter, people who appreciate self-deprecating humor, Kroger's occasionally nematode-free mushrooms, the NIH official who agreed that Reviewer 3 was wrong and gave me the F31 anyway, constructive reviews of scientific work, and people who strive to improve work-life balance and equity in academia.

My work has been generously supported by Vanderbilt University, the Gisela Mosig Travel Fund, and the National Institutes of Health.

Table of Contents

Page

Dedication ii
Acknowledgements iii
List of Tables ix
List of Figures x
List of Abbreviations xii

Chapter

Chapter Previews 1

1. Microorganisms in the reproductive tissues of arthropods 3
    Abstract 3
    Introduction 3
    Arthropod Reproductive Tissue Microbiota 4
        Transmission routes 7
        Box 1. Sexually transmitted infections of arthropods 13
    Evolutionary Impacts 14
        Impact on host fitness, development and ecology 14
        Gene expression, gene flow and genome evolution 16
        Host behavior, sexual selection and speciation 20
    Conclusions and Outlook 22
        Box 2: Microorganisms of human reproductive tissues 24
        Box 3: Microorganisms of plant reproductive tissues 25
    Acknowledgements 27
    Author Contributions 27

2. Microbe-mediated male killing 28
    Abstract 28
    Introdution 28
    Diversity of Male Killing 31
    Evolution of Host and Male-Killing Microbes 36
    Male-Killing Genetics and Mechanisms 41
    Applications of Male Killing 45
    Acknowledgements 46

3. Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility 47
    Abstract 47
    Results and Discussion 47
    Materials and Methods 60
        Comparative genomics and transcriptomics 60
        Protein extraction and mass spectrometry 61
        Evolutionary analyses 61

Gene expression assays and *Wolbachia* titers 62
Fly rearing 63
Transgenic flies 63
Hatch rate and sex ratio assays 64
Embryo cytology 65
Statistical analyses 66
Data availability 66
Acknowledgments 66
Author Contributions 67

4. The phage gene *wmk* is a candidate for male killing by a bacterial endosymbiont 68
Abstract 68
Summary 68
Introduction 69
Results 71
Genomic analysis of male-killing gene candidates 71
The *wmk* gene is common and found in all sequenced male-killing genomes 72
Transgenic expression of *wmk* causes a female-biased sex ratio 74
Transgenic expression of *wmk* recapitulates embryonic death and cytological defects 76
wmk is expressed in *Drosophila* embryos infected with *Wolbachia* 83
The Wmk protein is a putative DNA-binding protein 84
Discussion 85
Materials and Methods 90
Experimental design 90
Comparative genomics and evolutionary analysis 90
Wolbachia gene sequencing 91
Drosophila strains 92
Drosophila rearing 93
Sex ratio measurements 93
Hatch rates 94
Embryo cytology 94
Gene expression 97
Protein conservation 98
Statistical analysis 98
Data availability 98
Acknowledgements 98

5. Transgenic testing does not support a role for additional genes in male killing or cytoplasmic incompatibility 99
Abstract 99
Importance 99
Introduction 100
Results 102
Many additional *w*Mel genes are candidates for reproductive parasitism 102

The *Sp*AID homolog WD0633 in prophage WO does not transgenically recapitulate male killing or CI 104
Divergent *wmk* homologs in *w*Mel do not transgenically recapitulate male killing 107
Additional male-killing gene candidates identified via comparative genomics do not recapitulate male killing 109
wmk and *cifA* do not transgenically function together in reproductive parasitism 111
Transgenic expression with alternative transcripts of *wmk* results in loss of sex ratio phenotype 113
Discussion 115
Materials and Methods 122
Fly strains and transgene constructs 122
Fly maintenance 123
Sex ratio assays 123
Hatch rate assays 124
Domain and motif analyses 125
T4SS motif identification 125
Alternative start codon identification 125
Phylogenetic analyses 126
Protein alignment 126
Statistical analyses 126
Acknowledgements 127

6. Assessment of transgenic homologs of a male-killing gene candidate 128
Abstract 128
Introduction 128
Results 129
Discussion 138
Materials and Methods 140
Drosophila strains and maintenance 140
Sex ratio assays 140
RNA secondary structures 141
Gene expression 141
Sequence alignment 142
Statistical analyses 142
Acknowledgements 142

7. Conclusions and future directions 143
Are There More Male-Killers in Nature? 143
Why Is Male Killing So Taxonomically Diverse Compared to Other Manipulations? 143
How Do the Male-Killer and Host Evolve Over Time? 144
Are There More Male-Killing Genes in Nature? 145
What Are the Mechanisms of Male Killing and How Does Resistance Arise? 146
What is the Level of Specificity Between a Male-Killer and Its Host? 147
How do Male-Killing Symbionts Interact with the Rest of the Microbiome? 147

References 148

Appendix

A. Chapter III Supplementary Information 177
B. Chapter IV Supplementary Information 207
C. Chapter V Supplementary Information 224
D. Chapter VI Supplementary Information 233
E. Microbial misandry: Discovery of a *Spiroplasma* male-killing toxin 234
Summary 234
Main Text 234
F. List of Publications 238

List of Tables

Table **Page**

2-1. Reported cases of microbial male killing. 32
2-2. Cases of likely or known host resistance to male killing. 37
4-1. Comparative genomic analysis of male-killing gene candidates..... 72
5-1. List of gene candidates for male killing. 103
A-1. Core CI genome. This table shows the 161 unique *w*Mel genes identified as the "Core CI Genome" in Figure 1a. 189
A-2. Genes divergent in *w*Au. *w*Mel genes that are absent or divergent in *w*Au as identified in Fig. 1b. 194
A-3. *w*VitA transcriptome. *w*Mel homologs that are expressed (as measured by RNA) in the ovaries of *w*VitA-infected *Nasonia vitripennis*. 196
A-4. *w*Pip proteome. *w*Mel homologs detected at the protein level in *w*Pip (Buckeye)-infected *C. pipiens* ovaries. 201
A-5. Accession numbers. Accession numbers for WD0631/WD0632 homologs analyzed in Figs. 1c and 1e and Figs. A-1 to A-3. 203
A-6. Primers. Primers utilized in this study are listed. F = forward primer, R = reverse primer. 203
A-7. Exact p-values. The exact p-values for all statistical calculations, along with method used, are listed. 204
B-1. Full details of comparative genomics analysis for male-killing gene candidates..... 216
B-2. Homologs of Wmk from related bacterial strains. 219
B-3. Primers used in this study. 220
B-4. Wmk protein homologs included in Fig. B-1b. 221
B-5. *wmk* gene homologs included in Fig. B-1a. 221
B-6. Amino acid similarity between Wmk and homologs in male-killing strains..... 223
C-1. List of *w*Mel genes scored by S4TE algorithm. 227
C-2. List of alternative start codons to *wmk* and homologs. 231
D-1. Primers used in this study. 233

List of Figures

Figure **Page**

1-1. Example microorganisms associated with arthropod reproductive tissues. 5
1-2.Transmission routes of microorganisms in the reproductive tract in arthropods. 8
1-3. Effects on the genomes and transcriptomes of hosts and microorganisms. 17
2-1. Hand-drawn wing pattern of *Hypolimnas inoptina* W'house by Hubert Simmonds. 29
3-1. Comparative analyses reveal WD0631 and WD0632 in the eukaryotic association module of prophage WO as candidate CI genes. 49
3-2. Relative expression of CI candidate and prophage WO genes decreases as males age. 52
3-3. Dual expression of WD0631 (*cifA*) and WD0632 (*cifB*) is necessary to induce CI-like defects. 55
3-4. Dual expression of WD0631 (*cifA*) and WD0632 (*cifB*) recapitulates CI-associated embryonic defects. 59
4-1. Comparative genomics of *wmk* and its homologs in *w*Mel and male-killing strains. 73
4-2. Transgenic expression of *wmk* causes a female-biased sex ratio. 75
4-3. Transgenic expression of *wmk* causes cytological defects in early embryogenesis. 77
4-4. *wmk*-induced embryonic defects are enriched in males. 80
4-5. Transgenic expression of *wmk* causes DNA damage in association with H4K16ac. 83
4-6. Native *Wolbachia* gene and transgene expression in embryos of *D. melanogaster* and *D. bifasciata*. 84
5-1. Map of gene candidates assessed for reproductive parasitism across the *w*Mel genome.... 104
5-2. Transgenic expression of WD0633 does not recapitulate male killing or CI. 105
5-3. Transgenic expression of *w*Mel *wmk* homologs does not recapitulate male killing. 109
5-4. Additional male-killing gene candidates does not induce a biased sex ratio with transgenic expression. 110
5-5. Co-expression of *cifA* and *wmk* neither enhances the *wmk* sex ratio bias nor recapitulates CI induction. 112
5-6. Expression of *wmk* with an alternative upstream start codon results in loss of a sex ratio bias. 114
6-1. *wmk* homologs from male-killing strains distantly-related to *w*Mel do not induce a biased sex ratio in *D. melanogaster*. 130
6-2. *wmk* homologs most similar to *w*Mel *wmk* exhibit a range of transgenic phenotypes in *D. melanogaster*. 132
6-3. Predicted RNA secondary structures of the indicated native *Wolbachia* or transgenes exhibit large differences in structure. 133
6-4. *w*Rec and *w*Suzi transgenes expressed with an alternative start codon lose their transgenic phenotypes. 135
6-5. Changing the codon, but not the encoded amino acid, of the thirteenth amino acid position in *wmk* significantly alters resulting phenotype. 137
A-1. CI and the evolution of *Wolbachia* and prophage WO genes. 178
A-2. WD0631/WD0632 homologs always associate with the eukaryotic association module in prophage WO regions. 179
A-3. *Wolbachia* CI patterns correlate with WD0631/WD0632 homolog similarity and copy number. 180
A-4. *Wolbachia* titers, the male age effect, and the younger brother effect. 181

A-5. WD0625 transgene expression does not induce CI-like defects. 182
A-6. Expression of transgenes does not alter sex ratios..... 184
A-7. Transgenes are expressed in testes. 185
A-8. Transgenic expression of WD0508, WD0625, and WD0625/WD0632 does not enhance or induce CI. 187
A-9. Transgenic expression of control genes does not affect sex ratios. 189
A-10. There is variation in *Wolbachia* titers in transgenic lines..... 189
B-1. Comparative genomics of the *wmk* gene and protein. 207
B-2. *Act5c* and MTD expression of transgenes other than *wmk* do not cause a sex ratio bias. .. 208
B-3. The *wmk* phenotype can be induced with both the *arm* and *Act5c* drivers despite differences in expression levels. 209
B-4. The *wmk* phenotype is not due to other forms of reproductive parasitism and it does not induce or rescue CI. 212
B-5. The corresponding sex ratios of all experiments are female-biased. 213
B-6. Representative images of FISH staining of Y chromosome from data in Fig. 4-3. 214
B-7. Control gene females do not show DNA damage. 214
B-8. Predicted protein architecture of Wmk homologs and homology to a phage repressor. 215
B-9. Wmk amino acid identity is more conserved in the DNA-binding domains than certain other regions of the protein. 216
C-1. *Sp*AID is similar in sequence and structure to *Wolbachia* proteins. 224
C-2. Type IV secretion system (T4SS) motifs in WD0626. 225
C-3. Alternative transcripts of *wmk* may have different structures and functions. 226

List of Abbreviations

aa	amino acid
act5c	*actin5c* gene
AED	After Embryo Deposition
AICc	Akaike Information Criteria (corrected)
ANOVA	Analysis of Variance
arm	*armadillo*
BLAST	Basic Local Alignment Search Tool
BLASTn	Basic Local Alignment Search Tool (for nucleic acid sequences)
BLASTp	Basic Local Alignment Search Tool (for protein sequences)
cDNA	complementary Deoxyribonucleic Acid
CI	Cytoplasmic Incompatibility
cifA	*cytoplasmic incompatibility factor A*
cifB	*cytoplasmic incompatibility factor B*
CISR	Cell Imaging Shared Resource
CMY	Cornmeal Molasses Yeast
cpREV	chloroplast Reverse Transcriptase model of evolution
Ct	Cycle threshold
CyO	Curly balancer chromosome
DNA	Deoxyribonucleic Acid
DUF	Domain of Unknown Function
EAM	Eukaryotic Association Module
FISH	Fluorescent In Situ Hybridization
ftsZ	*filamenting temperature-sensitive mutant Z*
Gal4	Galactose 4 transcription factor
gpW	*gene product W* (bacteriophage baseplate protein)
groEL	*groEL* chaperonin gene
GTI	Genital Tract Infection
GTR	Generalized Time-Reversal model of evolution
H4K16ac	Histone H4 acetylation at lysine 16
HA	Hemagglutinin
HTH	Helix-Turn-Helix
Hz-2V	*Helicoverpa zea*-2 Virus (aka, Gonad-Specific Virus)
IS	Insertion Sequence
iTOL	interactive Tree of Life
JAMM	JAB1/MPN/Mov34 domain metalloprotease
JTT	Jones-Taylor-Thornton + Gamma model of evolution
mtDNA	Mitochondrial Deoxyribonucleid Acid
MSL	Male Specific Lethal

mRNA	messenger Ribonucleic Acid
MSRO	Melanogaster Sex Ratio Organism (*Spiroplasma poulsonii*)
MTD	Maternal Triple Driver
MUSCLE	Multiple Sequence Comparison by Log-Expectation
nos	*nanos* gene
NCBI	National Center for Biotechnology Information
NIH	National Institutes of Health
NSF	National Science Foundation
ORF	Open Reading Frame
OTU	Operational Taxonomic Unit
OTU Deubiquitinase	Ovarian Tumor domain deubiquitinase
PAZ	Piwi-Argonaut-Zwille domain of gene silencers
PBS	Phosphate-Buffered Saline
PBTA	PBS/Bovine Serum Albumin/Triton/Azide buffer
PCR	Polymerase Chain Reaction
PDDEXK	Motif for superfamily of nucleases
PFAM	Protein Family
pH2Av	phosphorylated histone H2Av
PLCV	Potato Leaf Curl Virus
PRIDE	Proteomics Identification Database-EMBL
pTIGER	plasmid for Targeted Integration Germline Expression UAS Regulated
RDV	Rice Dwarf Virus
RNA	Ribonucleic Acid
ROUT	Regression and Outlier removal
rp49	*ribosomal protein 49*
RPN	Recombination Promoting Nuclease
rRNA	ribosomal Ribonucleic Acid
RT-qPCR	Real Time quantitative Polymerase Chain Reaction
S4TE	Searching Algorithm for Type IV Secrection Effectors
SAP54	Secreted Aster Yellows Witches' Broom Phytoplasma Effector 54
SAP11	Secreted Aster Yellows Witches' Broom Phytoplasma Effector 11
SIT	Sterile Insect Technique
SMART	Simple Modular Architecture Research Tool
Spaid	*Spiroplasma* androicidin toxin
SSCT	Saline-Sodium Citrate Tween
STI	Sexually Transmitted Infection
T4SS	Type IV Secretion System
TCA	Trichloroacetic Acid
TF	Transcription Factor
TM	Transmembrane

TYLCV	Tomato Yellow Leaf Curl Virus
UAS	Upstream Activation Site
Ulp1	Ubiquitin-like-specific protease 1
VUMC	Vanderbilt University Medical Center
*w*AlbA	*Wolbachia* strain A of *Aedes Albopictus* mosquitoes
*w*AlbB	*Wolbachia* strain B of *Aedes Albopictus* mosquitoes
*w*Au	*Wolbachia Drosophila simulans* flies (Australian strain)
*w*Bif	*Wolbachia* of *Drosophila bifasciata* flies
*w*Bm	*Wolbachia* of *Brugia malayi* nematodes
*w*Bol1b	*Wolbachia* of *Hypolimnas bolina* butterflies
*w*Bor	*Wolbachia* of *Drosophila borealis* flies
*w*Caub	*Wolbachia* of *Cadra cautella* moths
wgs	whole genome shotgun
*w*Ha	*Wolbachia* of *Drosophila simulans* flies (Hawaiian strain)
*w*Inn	*Wolbachia* of *Drosophila innubila* flies
*w*Mel	*Wolbachia* of *Drosophila melanogaster* flies
*w*No	*Wolbachia* of *Drosophila simulans* flies (Noumean strain)
*w*Pip	*Wolbachia* of *Culex pipiens* mosquitoes
*w*Rec	*Wolbachia* of *Drosophila recens* flies
*w*Ri	*Wolbachia* of *Drosophila simulans* flies (Riverside, California strain)
wmk	WO-mediated killing
*w*Sim	*Wolbachia* of *Drosophila simulans* flies
*w*Suzi	*Wolbachia* of *Drosophila suzukii* flies
WT	Wild Type
*w*VitA	*Wolbachia* strain A of *Nasonia vitripennis* wasps
WO	*Wolbachia* phage
WOMel (etc.)	*Wolbachia* prophage of *w*Mel
WOMelB (etc.)	*Wolbachia* prophage region B of *w*Mel
XRE	Xenobiotic Response Element
YLS	Yeast-Like Symbiont

Chapter Previews

In this dissertation, I discuss my research on microorganisms of arthropod reproductive tissues. Chapter 1 provides a review of the principles of microbe-host symbiosis in the context of arthropod reproductive tissues. A comprehensive overview is given of microbial taxa present in the reproductive tissues of various arthropod hosts, as well as the dynamics of their transmission and interactions. I then describe their impact on the evolution of both host and microbe by describing the specific impacts microbial symbionts have on host physiology, fitness, genetics, behavior, and speciation, among other topics. Throughout these sections, I highlight important questions for future research and identify major gaps remaining in the field.

Building on this foundation establishing the importance of these symbioses, I then describe the current state of knowledge of male killing in Chapter 2. Specifically, I cover the diversity of male-killing symbioses in terms of both host and microbial taxa. I then describe the evolutionary impact of symbioses between male-killers and hosts including the potential outcomes for the microbes, as well as known and theoretical effects on hosts such as altered evolution of sex determination systems, reduced mtDNA haplotype diversity, and altered sexual behaviors leading to changes in sexual selection. I then cover what is known on the mechanism and genetics of male killing, describing known phenotypes as well as genes and probable mechanisms in various systems. This includes coverage of the work presented in subsequent chapters.

In Chapter 3, I present the publication that described the first comparative genomics analysis and functional testing of reproductive parasitism gene candidates in the lab. The candidates presented here include those that were identified as being likely candidates for the cytoplasmic incompatibility (CI) phenotype. We also established the pTIGER-based method of testing *Wolbachia* genes for reproductive parasitism. Although not discussed in this paper, the male-killing gene candidate my subsequent research focuses on was first identified in this paper as part of the "core" CI genome that led to its study in my later work. Thus, this paper sets the foundation for the identification and functional interrogation of *Wolbachia* reproductive parasitism gene candidates in *Drosophila melanogaster* and provided the first line of evidence for wmk as a manipulator of host reproduction.

Chapter 4 presents the main finding of my dissertation that wmk is the primary candidate for male-killing by *Wolbachia* in *Drosophila* hosts. I describe through comparative genomics and

other lines of evidence how the *wmk* candidate gene was first identified and then functionally test it for induction of a female-biased sex ratio in flies using the system established in Chapter 3. I then demonstrate recapitulation of several embryonic defects caused by natural male-killing infection, and also present evidence for a potential mechanism via interference with host dosage compensation. Additional evidence is also presented demonstrating that transgenic expression of *wmk* recapitulates many aspects of male killing, but not other forms of reproductive parasitism, and that it is not an artifact of transgenic expression.

In Chapter 5, I then transgenically test additional candidates for male killing in *D. melanogaster* and find no evidence that any other candidate identified thus far is involved in reproductive parasitism. In particular, I functionally assess candidates identified via comparative genomics in Chapter 4 in addition to other gene candidates for induction of a female-biased sex ratio. I present evidence that no gene candidate induces a biased sex ratio, making *wmk* the only one thus far that is able to do so. In addition, I find no evidence that wmk and the CI-causing genes function together. I also present evidence for a hypothesis that alternative *wmk* expression profiles may underlie differences between *Wolbachia* genotype and phenotype.

In Chapter 6, I describe the experimental results of transgenic expression of *wmk* homologs in *D. melanogaster*. Specifically, homologs of *wmk* from various *Wolbachia* strains in other hosts were transgenically expressed in the *D. melanogaster* background and tested for ability of these homologs to induce male killing. Although most homologs failed to recapitulate male killing in *D. melanogaster*, several exhibited unexpected phenotypes including death of males and females. I then discuss the possible causes for this and propose questions and experiments to be done to further assess the proposed hypotheses.

Chapter 7 then presents conclusions and future directions in the field of male killing. Specifically, I address potential applications of male killing to pest and vector control as well as conservation. I then end by presenting the important questions remaining in the field and provide discussion on each point regarding what is known and what should be done next.

Appendices A-D include supplemental information for Chapters 3, 4, 5, and 6. Appendix E presents an overview of a recent related publication describing the identification of a *Spiroplasma* male-killing toxin. Appendix F presents a list of publications.

Chapter 1

Microorganisms in the reproductive tissues of arthropods

Contributing Authors: Jessamyn I. Perlmutter and Seth R. Bordenstein

Abstract

Microorganisms that reside within or transmit through arthropod reproductive tissues have profound impacts on host reproduction, health and evolution. In this Review, we discuss select principles of the biology of microorganisms in arthropod reproductive tissues, including bacteria, viruses, protists and fungi. We review models of specific symbionts, routes of transmission, and physiological and evolutionary outcomes of both hosts and microorganisms. We also identify areas in need of continuing research to answer fundamental questions remaining in fields within and beyond arthropod-microorganism associations. New opportunities for research in this area will drive a broader understanding of major concepts, biodiversity, mechanisms, and translational applications of microorganisms that interact with host reproductive tissues.

Introduction

In 1879, Heinrich Anton de Bary, a German microbiologist and botanist, coined the term symbiosis to mean the living together of dissimilar organisms[1,2]. He devised the word symbiosis for his now famous talk discussing the relationships between algae, cyanobacteria and fungi that together form lichens[1]. Today, the term symbiosis generally describes any relationship type between or among different organisms, including mutualism (all parties benefit), commensalism (one party benefits while the other is unaffected) and parasitism (one party benefits while another is harmed). These relationships are often context-dependent, and additional categories or subcategories exist that are not necessarily mutually exclusive, such as endosymbiosis, whereby one organism lives inside another, and hereditary symbiosis, in which microorganisms are transmitted from parent to offspring[3-5]. Indeed, endosymbiosis and hereditary symbioses are common symbiotic relationships now recognized in many plants and animals[5]. Microorganisms of the reproductive tissues (discussed in this Review as those that reside in or transmit through reproductive tissues), including gonads, gametes and milk organs, are acquired from many different sources, including the environment and other host organisms (horizontal transmission) or from parent to offspring (vertical transmission). These microorganisms in particular can be key

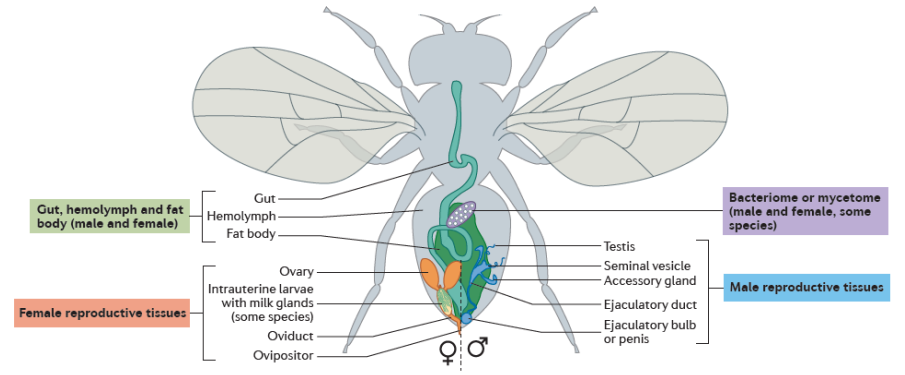
determinants of host fitness and offspring health because of their location and potential to be passed vertically. As microorganisms of the reproductive system are uniquely situated to alter host germlines and reproductive ability, it is crucial to understand their modes of transmission, functional relevance in hosts and effects on host evolution. Indeed, bacterial symbionts of arthropods are known to profoundly influence host reproductive strategies and physiology in ways that are often unique in the animal kingdom.

In this Review, we synthesize current knowledge on microbial symbionts that inhabit or transmit through the reproductive tissues of arthropods. We discuss which microorganisms are most often reported in these tissues, their various modes of transmission, and the influence of these symbioses on the evolution of hosts and microorganisms. We also assess widespread and specialized biological principles across various organisms and highlight major fundamental, unanswered questions in need of continued study. We then emphasize important future directions in the field, including a call for more microbial community sequencing in reproductive tissues, an increased focus on non-bacterial members of the microbiota and greater study into microorganism-microorganism interactions in reproductive tissues. New discoveries in this arena will spur innovation and discovery in both the basic and applied sciences, including vector and pest control efforts and a greater understanding of the impact of microorganisms-associated with the reproductive tract on host evolution.

Arthropod Reproductive Tissue Microbiota

Microorganisms that inhabit arthropod reproductive tissues represent an exceedingly broad group of organisms spanning many orders of bacteria[6-14], fungi[15-18], protists[16,19] and viruses[16,20-22] (Fig. 1). Microorganisms are present in the reproductive tissues of all of the major orders of arthropods[5], including various insects[15,23], crustaceans[24,25] and arachnids[26,27] around the world. Spanning the entire range of symbiotic relationships with their hosts, these microorganisms vary from transient pathogens to obligate mutualists, and they perform various functions within hosts. Among these are the well-known bacterial, viral and fungal reproductive parasites that manipulate host reproduction[6,20,23,28-30]; bacteria, viruses, fungi and protists that cause sexually transmitted diseases in their hosts[16]; commensal or harmful bacteria and viruses that use arthropods as vehicles to infect plants or other animals[22,31-33]; bacterial nutritional symbionts that provide essential vitamins and other nutrients to the host[8,13,34]; bacteria and viruses that protect hosts from

predation[10,35-38]; and bacteria that perform nitrogen cycling for the host[13,39], among many others. Owing to their successful survival strategies in the most speciose groups of animals, microbial inhabitants of arthropod reproductive tissues represent some of the most widespread symbioses in nature.



	Gut, hemolymph and fat body (male and female)	Female reproductive tissues	Male reproductive tissue	Bacteriome or mycetome (male and female, some species)
Bacteria	*Borrelia* (ticks, lice) *Francisella* (arthropods) *Burkholderia* (bean and stink bugs) *Sodalis* (tsetse flies) *Stammera* (tortoise leaf beetles) *Pseudomonas* (beetles) *Asaia* (mosquitoes) *Candidatus* Ishikawaella (kudzu bugs) *Blattabacterium* (cockroaches) *Blochmannia* (ants) *Rhodococcus* (reduviid bugs) *Spiroplasma* (flies)	*Ehrlichia* (ticks) *Midichloria* (ticks) *Orientia* (ticks) *Coxiella* (ticks) *Nasuia* (leafhoppers) *Sodalis* (tsetse flies) *Stammera* (tortoise leaf beetles) *Pseudomonas* (beetles) *Asaia* (mosquitoes) *Coxiella*-like symbionts *Ehrlichia*-like symbionts *Francisella*-like symbionts *Rickettsia*-like symbionts *Morganella* (grass grub) *Blochmannia* (ants) *Cardinium* (arthropods) *Arsenophonus* (arthropods) *Rickettsia* (arthropods) *Spiroplasma* (arthropods) *Wigglesworthia* (tsetse flies) *Wolbachia* (arthropods)	*Sodalis* (tsetse flies) *Asaia* (mosquitoes) *Cardinium* (arthropods) *Wigglesworthia* (tsetse flies) *Arsenophonus* (arthropods) *Rickettsia* (arthropods) *Spiroplasma* (arthropods) *Wolbachia* (arthropods)	*Baumannia* (sharpshooters) *Buchnera* (aphids) *Carsonella* (psyllids) *Moranella* (mealybugs) *Nardonella* (weevils) *Nasuia* (arthropods) *Portiera* (whiteflies) *Riesia* (lice) *Sulcia* (arthropods) *Tremblaya* (mealybugs) *Uzinura* (scales) *Hamiltonella* (aphids and other sap feeders) *Regiella* (aphids) *Serratia* (aphids) *Wigglesworthia* (tsetse flies)
Fungi	*Candida* (arthropods) Yeast-like symbionts (YLS, arthropods) Microsporidia (arthropods) *Massospora* (cicadas)	*Candida* (arthropods) Yeast-like symbionts (YLS, arthropods) Microsporidia (arthropods)	Yeast-like symbionts (YLS, arthropods) Microsporidia (arthropods)	Yeast-like symbionts (YLS, arthropods)
Viruses	Arboviruses (mosquitoes, ticks, other vectors) Phage (bacteria of arthropods)	*gypsy* retrovirus (flies) Iflaviruses (arthropods) Nyamiviruses (ticks) Rice dwarf virus (leafhoppers) Sigma viruses (dipterans) Phage APSE (*Arsenophonus* of arthropods) Begomoviruses (whiteflies) Polydnaviruses (wasps) Phage WO (*Wolbachia* of arthropods) Gonad-specific or Hz-2V virus (moths)	*gypsy* retrovirus (flies) Iflaviruses (arthropods) Nyamiviruses (ticks) Sigma viruses (dipterans) Phage APSE (*Arsenophonus* of arthropods) Phage WO (*Wolbachia* of arthropods) Gonad-specific or Hz-2V virus (moth)	Phage APSE (*Hamiltonella* of aphids and other sap feeders)
Protists	*Babesia* (ticks)	*Babesia* (ticks)		

Parasite or pathogen
Reproductive parasite
Nutritional and/or nitrogen recycling symbiont
Defense against parasitoids or predators
Mating behavior
Domesticated for parasitization of secondary host
Contributes to host development or reproduction

Figure 1-1. Example microorganisms associated with arthropod reproductive tissues. The silhouette is a representative image that includes organs from both males and females, as well as various species of insects, and is therefore not anatomically accurate for any given arthropod species, and neither is it to scale to enable visualization of all organs. Select microorganisms and viruses (the general category or the genus level is listed, if known) are listed in their primary or additional densely populated body sites. The list is not comprehensive of all symbionts, all tissue localizations or all functions, but it represents many known symbioses. In addition, symbionts may

not be present in the same tissues or exhibit the same phenotypes with every microorganism strain or host. The microorganisms listed under 'Gut, haemolymph and fat body' or 'Bacteriome or mycetome' are present in both sexes of some species and are included because all these microorganisms contact reproductive tissues at some point (typically during transmission), even though they primarily or often reside in somatic tissues.

Though a great number of specific microorganisms and symbiotic relationships exist across the range of arthropod hosts, little is known about the complete diversity of microorganisms or microbial community interactions within host reproductive tracts with notable exceptions[9,40-43]. Indeed, binary microorganism-host symbioses, particularly those that are hereditary, represent the majority of published research in the field. One of the few microbial community characterizations thus far was completed in the reproductive tract of two *Anopheles* mosquito species. The study reported that these mosquitoes contain on average 500 species-level bacterial OTUs in the reproductive tissues, and that there is a core microbiota spanning seven genera shared among individuals of the same sex and swarm[9]. Another study showed that bedbugs exhibit a diversity of bacteria in their reproductive organs, with 31 sequence variants found across samples, although individuals harbor an average of just three sequence variants[44]. In addition, differences in communities occur between males and females and mated and unmated individuals, which suggests sex-specific differences and exchanges of microorganisms during copulation[44]. In both studies, sequencing was not performed on contamination controls, and staining was not performed to confirm the presence of live bacteria. However, the data suggest that there may be important factors such as sex, proximity or relatedness of individuals that correlate with community structure. It also cautiously raises the potential that reproductive tissues of some arthropods may harbor a diversity of microorganisms. In addition, there are studies that focus on the interactions of several select symbiont species or strains within one host[11,42,45]. Many of these investigations survey infection and co-occurrence rates in a population and illustrate the potential positive or negative influences that specific bacterial reproductive parasites have on each other's transmission[40,46,47]. However, studies simultaneously assessing multiple microorganisms within arthropod reproductive tissues remain relatively scarce, and this is especially true for whole-community analyses.

We have little knowledge of how diverse the microbial communities are within arthropod host reproductive tissues, their temporal dynamics, how much community compositions vary between individuals or species, how well they correlate with host species phylogeny, what factors

or conditions shape microbial communities, what kinds of interactions exist across the microbial community or how they are acquired. In addition, surveys of non-bacterial taxa within arthropods are underrepresented in the literature, including those of viruses, fungi, protists and archaea. In particular and to the best of our knowledge, there are no studies that have investigated archaeal symbionts of arthropod reproductive tissues. Although archaeal methanogens are found in guts of termites and other insects[48], potential archaeal roles in reproductive tissues are largely unknown. Furthermore, research is far more common in insects than arachnids, crustaceans and other arthropods, although there are a few studies describing unidentified bacteria in the reproductive tissues of animals such as shrimp[49] and crabs[50], and some that identify endosymbionts of arachnids such as spiders and mites[20,51-54]. Given the growing recognition of the role archaea have in the health of organisms, including humans[55], and an increased understanding of the role of microbial community dynamics in the functions of diverse hosts[56,57], these are important research frontiers for the field to explore. Therefore, it will be crucial to better characterize the identities and dynamics of all members of the reproductive tissue microbiota of many hosts, as well as to emphasize additional research on symbiotic interactions in the context of their community rather than only in an isolated host-microorganisms relationship.

Transmission routes

Transmission routes of microorganisms within arthropod reproductive tissues have been extensively explored[5,58]. The various transmission routes can be compared along numerous axes, including horizontal and vertical, maternal and paternal, intracellular and extracellular, sexual and non-sexual, host-driven and microorganism-driven, or transmission driven by microorganism-microorganism interactions and microorganism-host interactions, and these categories are not mutually exclusive (Fig. 2). As hereditary microorganisms are arguably the best studied within the arthropod reproductive tract, a large portion of the research in this area has focused on vertical transmission routes.

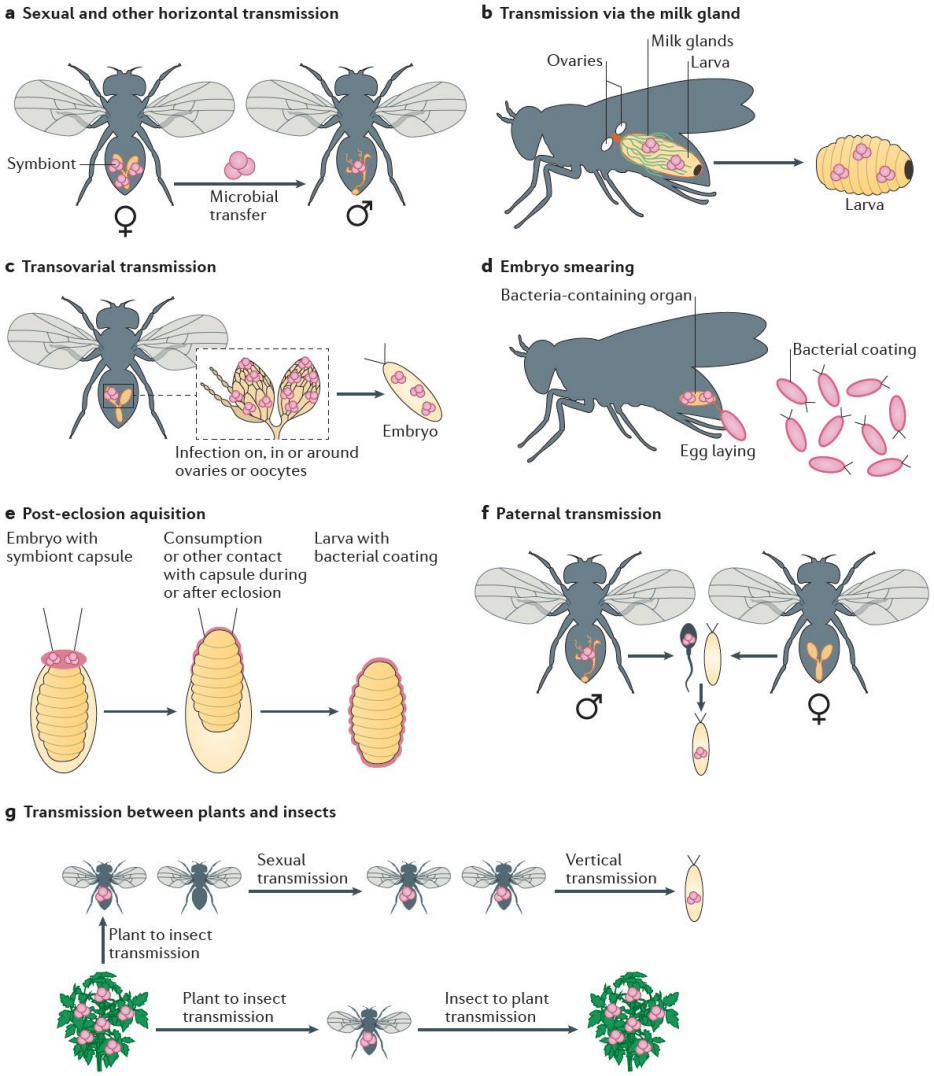


Figure 1-2.Transmission routes of microorganisms in the reproductive tract in arthropods. Depicted are representative methods for the transmission of microorganisms in the reproductive tract between individuals. Pink circles represent symbionts, and pink outlines or coatings indicate outer coverage by symbionts. a) Horizontal transmission can spread microorganisms between the reproductive tissues of different host individuals, usually through copulation. b) Certain hereditary microorganisms can be vertically transmitted from mother to offspring via the milk glands, as has been reported for the tsetse fly symbiont Wigglesworthia glossinidia. c) Hereditary endosymbionts, including many common reproductive parasites, can be vertically transmitted via infection in the ovarian tissues and passage internally to embryos. d) Bacteria in specialized organs can be smeared onto embryos as they are laid so that offspring are coated with the microorganisms when they eclose. e) Hereditary symbionts may be acquired post egg laying through various mechanisms, including passage through symbiont capsules during eclosion. f) Microorganisms may be paternally transmitted via various mechanisms, including packaging within sperm heads that enables infection of the offspring. g) Certain microorganisms may also be cyclically transmitted through both insect and plant hosts. The insects often carry these microorganisms on their genitalia and can pass them sexually to other insects, horizontally to new plants or vertically to offspring.

One common form of vertical inheritance is transovarial transmission (inside the egg), which is typical for hereditary endosymbionts such as *Wolbachia*[59], *Rickettsia*[60], *Spiroplasma*[61], *Buchnera*[62], and *Hamiltonella*[63] spp., as well as certain fungi such as yeast-like symbionts (YLS)[64] and microsporidia[65], some viruses[66] and protists[19]. There are various mechanisms to ensure this form of passage, which are often microorganisms-driven. For example, *Wolbachia pipientis*, hereafter referred to as *Wolbachia*, infects germline stem cells[67] to spread into oocytes using host actin during oogenesis[59]. They subsequently use egg microtubules to localize towards the posterior end of the embryo where cells are fated to become germline tissue[68], and the cycle repeats in the next generation. *Spiroplasma poulsonii* instead hijack the yolk uptake machinery to be endocytosed into the oocyte[61]. Alternatively, *Buchnera aphidicola* cells are carried by maternal bacteriocytes in the midgut and are exocytosed into the extracellular space for a short period of time before selective endocytosis into the maternal blastulae at ovariole tips[69]. Indeed, internally transmitted parasitic microorganisms that hijack animal reproductive processes (for example, *Wolbachia* and *Spiroplasma* species) often drive their own transmission, whereas beneficial symbionts that are vertically inherited (for example, *Buchnera aphidicola*) often rely on host-driven processes that ensure passage to the next generation. In many cases in which the microorganism is vertically transmitted, the exact molecular details are not fully understood. Therefore, it will be important to continue to interrogate the molecular, genetic and biochemical mechanisms, especially among non-bacterial symbionts that have not been studied as extensively.

External vertical transmission (outside the egg) is also a typical route of transmission for bacteria and fungi, and there are many variations on this theme in nature[70]. One common mechanism is smearing the symbiont on the egg as it exits the ovipositor. For example in the tortoise beetle, the obligate, beneficial *Stammera* bacterial symbiont that is essential for breaking down pectin in the host's plant-based diet is transmitted from specialized reservoirs connected to the ovipositor, so eggs are covered by the time they are laid[71,72]. This is similar to the transmission of microorganisms in vaginally delivered human babies[73]. However, there are many other modes of external transmission of symbionts that are known, particularly for bacteria. Sometimes the mother will produce secretions with beneficial microorganisms and deposit them onto eggs, which is the route of transmission of the bacterial symbiont *Candidatus* Tachikawaea gelatinosa in urostylidid stinkbugs[74]; or the mother will secrete the substance containing the symbiont onto the

surrounding area, as has been shown for necrophagous beetles[75]. In other cases, the mother might package the bacteria into capsules on egg cases that are eaten as larvae emerge[72], or secrete the obligate, beneficial bacterial endosymbionts in a milky substance for her developing offspring[12,76]. In addition, parasitic microorganisms may drive their own external transmission in some cases. For example, the male-killing endosymbiont *Arsenophonus* of *Nasonia* wasps exhibits a temporary tropism for the developing wasp oviduct and ovipositor, which promotes the external transfer of the bacteria via a transovum route to fly hosts. Subsequently, larval wasps feeding on the fly host become infected, and the transmission cycle repeats itself generation after generation[77]. Despite the ever-growing knowledge of unique mechanisms of external transmission, many questions remain. For example, what is the full diversity of external transmission modes in nature? What are the external transmission routes of understudied non-bacterial symbionts, and do they differ from bacterial symbionts?

Many studies focus on the influence of host-microorganism interactions on transmission, but there is also an emerging interest in the impact of microorganism-microorganism interactions on transmission. In particular, hereditary symbionts often have the unique position of being the first microorganisms in or on the next generation of offspring. Such founding microorganisms could have the potential to shape downstream microbial community assembly and composition via positive or negative interactions with other microorganisms and therefore, may be important determinants of offspring health. Microorganisms present in the parental reproductive tissues can notably affect which other microorganisms are passed to arthropod offspring. For example, newly acquired *Wolbachia* in *Anopheles* mosquitoes are not transmitted to the next generation[43] owing to negative interactions with the native microbiota. More specifically, if antibiotics are used to perturb existing members of the microbiota, *Wolbachia* are transmitted. If the resident gut and reproductive tissue bacteria, specifically of the *Asaia* genus, are supplemented back into the mosquitoes after antibiotic treatment, *Wolbachia* are no longer transmitted, which shows that bacteria in the *Asaia* genus negatively affect *Wolbachia* transmission[43]. Moreover, certain populations of pea aphids are infected with many different hereditary endosymbiont species[40]. Monitoring co-infection frequencies over time revealed that certain combinations of endosymbionts are more common than others. This suggests that microorganism-microorganism interactions within hosts have an impact on the transmission rates of these endosymbionts. In addition, microorganism-microorganism interactions can be costly or beneficial to the interacting

symbionts and thus may affect increases or decreases in the proportion of individuals containing multiple symbionts in a population over time[40]. Data support the notion that interactions among hereditary microorganisms can have lasting effects on which microorganisms get passed down to the next generation, which may have important implications for the health and fitness of the offspring. However, the fitness effects of these interactions have not all been fully experimentally explored. In addition, there has typically been a focus on how a select group of bacteria positively or negatively affects each other's transmission; however, it remains to be elucidated how a microorganism or group of microorganisms affect the larger ecosystem of the symbiont and to what extent in the host and offspring.

It is perhaps not surprising that interactions between microorganisms of different taxa affect each others' transmission, although it is a comparatively rare topic in the literature. In one intriguing case, the rice dwarf virus (RDV) is vectored between plants by leafhoppers via a unique type of vertical transmission[11]. The virus is associated with an insect obligate bacterial symbiont, *Sulcia,* and the virus hitchhikes on the envelope of the *Sulcia* symbiont via an interaction between a viral capsid protein and an outer membrane protein of *Sulcia*. After attachment, the virus uses *Sulcia* bacteria as a vehicle for transovarial transmission[11]. This case exquisitely highlights largely underappreciated interactions between microbes of different classifications of life that influence each other's inheritance. Indeed, the RDV-*Sulcia* interaction is probably not unique as additional studies have reported data demonstrating interactions that affect inter-taxa transmission. For example, transmission of tomato yellow leaf curl virus (TYLCV) vectored by whiteflies depends on the chaperone protein GroEL from *Hamiltonella*[78], and transmission of potato leafroll virus (PLRV) requires the aphid endosymbiont-specific protein symbionin for transmission[79]. Moreover, other studies reported the negative interactions of *Wolbachia* with the vertically inherited *gypsy* retrovirus[80] as well as Zika virus and other viruses in cases of hosts infected with non-native strains of *Wolbachia*[81,82]. Future investigations are required to determine how common these interactions are and between which taxa. Are there interactions that occur among or between fungi, protists or any potential archaea and other microorganisms? Are there cases whereby more than two entities directly interact in transmission? Do multipartite interactions occur in well-studied binary symbioses of host and microorganism? How would a microorganism evolve transmission dependency on another microorganism rather than the host? How would this partnership in transmission affect the evolution of each member of the system over time? Do these

interactions more easily develop with obligate symbionts than facultative microorganisms, or with parasites or mutualists? Multipartite transmission interactions represent an important future research area.

Maternal transmission is the most common form of transmission for bacteria rather than paternal transmission owing to the removal of cytoplasmic material during spermatogenesis; however, there are rare cases of paternal inheritance. For example, *Rickettsia* symbionts of leafhoppers are found within sperm heads and are transferred to offspring via the sperm[83]. In addition, bacterial endosymbionts of insects, such as those of the genus *Asaia* of *Anopheles* mosquitoes or *Sodalis* of tsetse flies, are sexually transmitted from males to females and subsequently passed on vertically to offspring[84,85]. Although rare for bacteria, many viral endosymbionts can also be both maternally and paternally transmitted, including the well-studied cases of sigma viruses in insects[86]. Sigma viruses are relatively common symbionts in insects, but they are unusual as they are one of the few known insect viruses that are exclusively vertically transmitted via gonads; the transmission from females is still much more efficient than males, probably owing to the lower number of viral particles that can be packaged in sperm compared to eggs[86]. Transmission by both parents enables symbionts to persist in additional contexts compared with symbionts with only one mode of transmission. Despite this benefit, there are far fewer known cases of paternal transmission of microorganisms in nature, even across different microbial taxa. This may either reflect actual rarity in nature or that paternal transmission is understudied in some contexts[87]. In addition, biparental transmission routes can result in important consequences that differ compared to organisms that are strictly or nearly always maternally transmitted. This includes whether or not male host fitness benefits the microorganism and the different infection rates and dynamics in a host population[88,89]. Future research could help identify additional cases of paternal transmission across microbial taxa and assess common and differing biological principles that might link to different transmission routes or microbial taxa.

The subfield of microbial transmission routes in arthropod hosts is vast, and many questions remain. In particular, although there is extensive work on vertical inheritance of single bacterial symbionts in arthropods, other categories are not as well represented. Beyond the relatively understudied non-microbial taxa previously mentioned, less is generally known about more transient microorganisms associated with arthropod reproductive tracts, excluding the many known sexually transmitted infections (STIs)[16] (Box 1), and there is a need for increased

investigation into opportunistic microorganisms of insect genitalia[90]. Moreover, how are microorganisms that are not vertically transmitted acquired, and are there any factors that select for certain microorganisms over others? Is there a 'core' or 'healthy' microbiota associated with arthropod reproductive tissues? What types of interactions exist between transient microorganisms and non-transient or inherited members of the microbiota? Those questions will need to be addressed in future research.

Box 1. Sexually transmitted infections of arthropods

Sexually transmitted infections **(**STIs) of arthropods are diverse and span many different bacteria, fungi and viruses and have been extensively reviewed elsewhere[16,90]. Some exhibit mixed modes of inheritance (that is they are both sexually and vertically transmitted[91,92]), so the two modes are not mutually exclusive. Despite the different form of transmission, vertically and sexually transmitted organisms share dependency on host reproductive activity and thus share many of the same biological principles related to host fitness and evolution. For example, STIs can affect host fitness by altering egg production rates[93] or sperm motility[94]. They may also sterilize the host[94,95], thus sharing the consequence of reduced offspring that is characteristic of symbionts that manipulate host reproduction. Beyond direct fitness impacts, interactions between hosts and STIs have led to reproduction-specific immune responses and defenses that are often unique to the reproductive tissues[96] or are specifically modulated to prevent STIs during mating activity[97]. These immune responses are a likely to be a result of an evolutionary arms race between STIs and their hosts. Indeed, antimicrobial peptides are commonly found on eggs[98,99] as well as seminal fluid[100] to help protect females and offspring from infection.

Beyond a direct fitness effect, STIs share some important host evolutionary consequences with vertically inherited symbionts. Indeed, STIs may induce altered host mating behaviors to facilitate their spread, such as viruses that correlate with quicker mating rates in males[94]. Importantly, they may also represent an environmental reservoir for the establishment of new host-microorganism symbioses. For example, one study demonstrated that aphids may acquire new beneficial symbionts initially through sexual transmission[92]. When experimentally tested, bacteria carried by male aphids could be sexually transferred to their female partners and subsequently transmitted vertically via the matriline. Further, they could replace other symbionts already carried by the mothers[92]. This suggests that some current day symbioses may have originally begun as

STIs. Therefore, STIs have much of the same potential to affect host evolution as the well-studied vertically inherited symbionts. However, there are some differences, including that STIs tend to be more often pathogenic, may have a wider host range, rely on host males and females for dispersal more equally and are horizontally rather than vertically transmitted[16,90].

Evolutionary Impacts

Impact on host fitness, development and ecology

Microorganisms that inhabit or transmit through arthropod host reproductive tissues can have a fundamental impact on host fitness and physiology. For example, they can damage or destroy reproductive tissues[16,17,101,102]; affect fecundity[103-105], oogenesis[106,107] or spermatogenesis[108]; have crucial roles in nutrient provisioning[39]; influence offspring development rate[103,109,110]; and affect predation or pathogen susceptibility[10,35,111-113]. These effects may also extend beyond a single generation and affect long-term physiological development, survival or evolution of the host. One interesting case is that of the rove beetle and its vertically transmitted *Pseudomonas* endosymbiont that produces the polyketide pederin, which protects the host from predators[38]. Over time hosts may even develop unique or specialized organs or proteins that function in housing symbionts (many are referred to as bacteriomes or mycetomes)[7], controlling their transmission[114] or preventing pathogen transmission during mating[97,115]. Another example is bedbugs that have a unique and costly form of copulation[116] whereby females are traumatically wounded during insemination. The male organ will pass through a specialized female organ, the spermalege, that has evolved at least in part to defend against pathogens that may be introduced during traumatic insemination[117]. Bedbugs also have a mycetome attached to the gonads that allows vertical transmission of symbionts[118,119]. As evidenced with bedbugs, specialized organs that defend against harmful symbionts and house helpful symbionts may evolve in the same host. Certain symbionts may even affect the size or shape of reproductive organs, such as with *Wolbachia*-infected crickets[120]. In some populations of crickets, *Wolbachia*-infected females have different spermathecae duct lengths than their uninfected counterparts, and this difference is recoverable following antibiotic treatment that removes the *Wolbachia* infection[120]. Therefore, symbionts can influence not only the long-term evolutionary development of arthropod organs but also individual reproductive organ physiology as well.

In addition to physiological and fitness effects on arthropods, microorganisms of reproductive tissues may directly affect host reproduction. In particular, reproductive parasites span diverse bacterial[121], fungal[122] and viral[123-125] lineages and selfishly manipulate host reproduction to facilitate their own spread at the expense of the host. Resulting phenotypes in the host can include cytoplasmic incompatibility (death of offspring from crosses between infected males and uninfected females), male killing (specific death of male offspring), feminization (genetic males physically develop and reproduce as females), masculinization (genetic females physically develop as males) and parthenogenesis (female reproduce asexually)[23,124]. Indeed, in long-term symbioses of this nature, or in cases of horizontal gene transfer from microorganism to host, the evolution of host reproduction, sex determination or sex development may be altered[126,127]. For example, long-term parthenogenesis in wasps can lead to an obligate dependency on the reproductive parasite owing to the accumulation of mutations in male-specific genes and phenotypic erosion of the ability to sexually reproduce[126,128]. In another case, feminizing *Wolbachia* of an isopod host are not transmitted via females with YY chromosomes, which resultantly produce all-male offspring, which is an intriguing case whereby a host allele may have evolved on sex chromosomes to avoid population-level extinction[129]. Microorganisms of reproductive tissues therefore specifically benefit from influencing reproductive features and behaviors of the host that would not necessarily be advantageous to microorganisms of other sites, and they do so in different ways. For example, some bacteria may benefit by manipulating reproduction to increase the fecundity of their transmitting hosts. However, fungi may benefit by slowly killing their transmitting host to enable many spores to develop and spread to new hosts[130]. Viruses may potentially do either, as they function as reproductive parasites that kill during either early or late developmental stages[28,131]. Although there are some extraordinary cases where evolution of arthropod sex development and determination are shaped by symbiosis, it is not known how common this phenomenon is nor are all of the mechanisms fully understood.

Beyond reproductive characters, microorganisms in the reproductive tracts can substantially affect the ability of their host to occupy a particular ecological niche. In some cases, the symbiosis affects the animals or plants with which the host interacts[132-134], and in others, endosymbionts confer differential temperature tolerance to their hosts that may narrow the range of environments suitable for the host[135] either by increasing[136] or decreasing[137] the range of tolerable temperatures to the host. In addition, environmental temperature can determine symbiont

phenotype owing to cold or heat sensitivity, as well as transmission. This has consequences for the spatial distribution of the host and symbiont and their ability to spread into new populations or survive in new environments[138]. However, what are the molecular mechanisms of these changes (known in some cases to be due to the induction of heat shock proteins by the symbiont[136])? How does niche specialization begin and develop over the course of a symbiotic relationship? Are the changes generally host-driven or microorganism-driven, and how do the changes differentially affect the fitness of the host and microorganism?

Gene expression, gene flow and genome evolution

Host-microorganism interactions in the reproductive tissues not only affect fecundity and sexual selection, but they also have a substantial impact on the genome and transcriptome through modulation of gene expression; interdomain transfer of genes between the interacting partners; and evolutionary pressures acting on different genomes that have intertwined fates (Fig. 1-3).

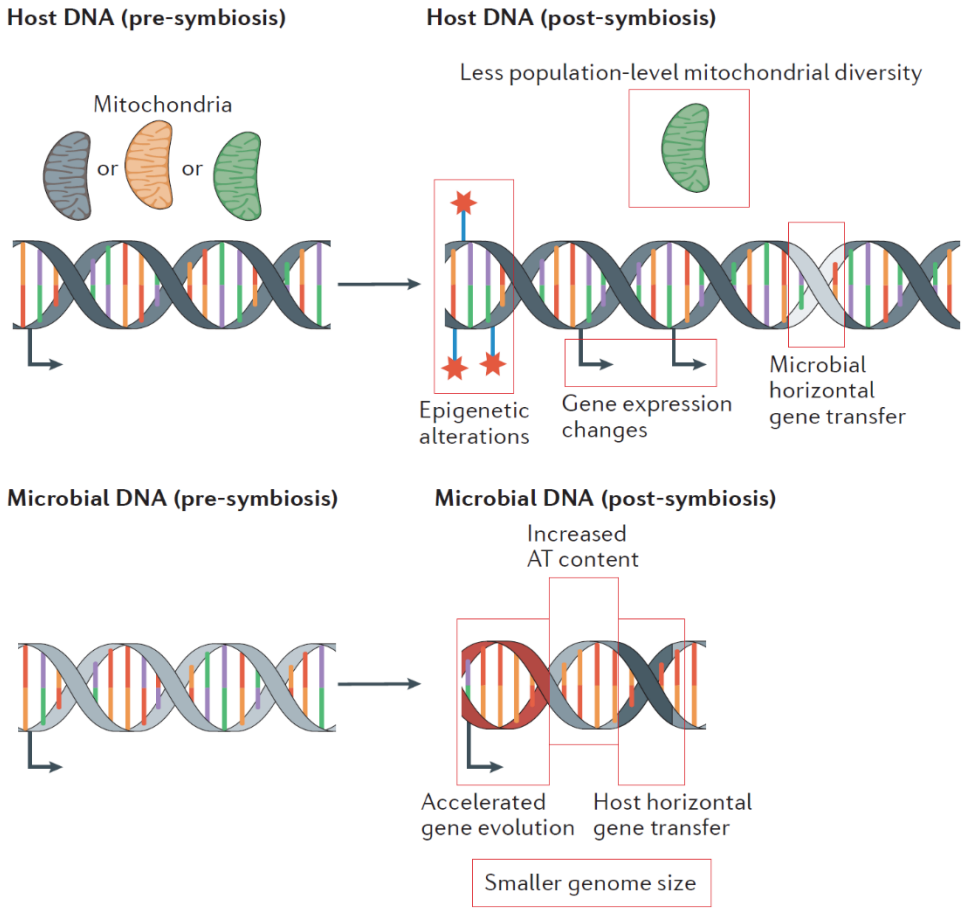


Figure 1-3. Effects on the genomes and transcriptomes of hosts and microorganisms. Each effect is not universal to all symbioses but instead represents changes known in some systems. The top panels show changes that can occur in host DNA, such as epigenetic alterations, changes in gene expression in the presence of symbionts, fewer mitochondrial DNA haplotypes in the population and horizontal gene transfer from microorganisms to host. The bottom panels show changes that can occur in microbial DNA, such as accelerated gene evolution, increased AT content, horizontal gene transfer from host to microorganism and an overall reduced genome size. The post-symbiosis panels show population-level changes many generations since pre-symbiosis.

Several studies demonstrate the effect of a microbial symbiont on host reproductive tissue gene expression and imprinting[115,139-142], with the noteworthy caveat that some amount of somatic tissue was sometimes pooled with reproductive tissues[139]. Potentially hundreds of genes are differentially regulated in reproductive tissues in the presence of a symbiont[139,140], gene expression profiles are different among infected soma and germline tissues[24], host genes may be differentially expressed in response to an endosymbiont in male and female tissues[139], and many of the genes that are differentially expressed have a role in metabolism, immunity and sex-specific developmental processes such as spermatogenesis[24,115,139-141,143]. The reciprocal analysis of

symbiont gene expression changes in soma and germline tissues is less common and thus results cannot yet be generalized. However, it is possible for symbiont genes to be expressed differentially in male and female reproductive tissues of a host[144]. Although the mechanisms underlying the transcriptional changes are not fully resolved, symbiont-mediated epigenetic changes in the host are common, particularly in parasitic relationships[142,145,146]. The role of microbially mediated epigenetic changes in host gene expression has mainly been studied in *Wolbachia*[145,146]. Other endosymbionts such as *Buchnera* lack genes for DNA methylation[147], so there may be differences in any putative regulatory mechanisms across organisms. This body of work has generated many questions to be more broadly explored in the future. How are all of these complex transcriptional responses and relationships regulated and mediated? In what circumstances are epigenetic changes in the host modulated, and are they changed directly or indirectly by symbionts? How do expression patterns differ across pathogens and mutualisms or dependencies? What are all of the transcriptional trends across more recently evolved symbioses compared to ancient relationships? Are the general principles different across microbial taxa (non-bacterial symbionts are currently underrepresented)? Although these questions have been answered in some cases, there is a further need to assess them in additional contexts and organisms.

Interactions between microorganisms of reproductive tissues and hosts also shape their genome content via horizontal gene transfer. Instances of gene exchange between hosts and microorganisms have attracted considerable attention because such exchanges can potentially bestow novel genes or larger sequences of DNA that are functional and/or heritable in the recipient genomes. Many such transfers have been reported for diverse microbial taxa[148]. Indeed, microorganism-to-host transfer events can include single genes[149-151], larger genomic regions[152] and nearly entire genomes[127,153], with some hosts containing genes from multiple symbionts[152]. For example, a single bacterial gene encoding cytolethal distending toxin was transferred to fly and aphid genomes and is likely to function in host defense[150], and insertion and duplication led to the presence of several megabases of *Wolbachia* DNA in the genome of the *Armadillidium pillbug* and may underpin the development of a new sex chromosome[127]. In addition, many of transferred genes confer functions such as detoxification[154] and thus host protection. An intriguing case of gene transfer is that of the *oskar* gene, which is present in many insects and is crucial for host germline development[155]. Part of the gene may have bacterial origins, thus making it a rare, putative case of bacterial gene transfer that functions in host reproduction[155]. Transfers may also

occur from host to microorganisms of various taxa[20,156,157], although fewer cases of this phenomenon have been discovered thus far. For example, prophage WO of *Wolbachia* harbors a eukaryotic association module that is composed of genes with regions of arthropod-like DNA[20]. This phage module is expressed within arthropod gonads and some genes of this module manipulate arthropod reproduction[158-161]. In light of the rarity of reported host-to-microorganism transfers, it is difficult to determine whether a particular function of transferred genes is enriched. It is likely that fewer eukaryotic genes are transferred to and maintained in bacteria due to inclusion of introns, exons and other elements that do not translate well in a bacterial genome, or it is possible that eukaryotic genes are retained less often due to their generally larger size. In addition to the above trends, it is apparent that most known transfer events occurred between the host and either bacteria or fungi, and there are relatively fewer examples known among viruses or other microorganisms[148]. With great progress in this area in recent years, new research questions are now at the forefront of the field. For example, is the rate of gene transfer and maintenance different between microorganisms that primarily or exclusively occupy reproductive tissues compared with microorganisms that primarily occupy soma? What are the relative rates of transfer in each direction between host and microorganism, and do these rates differ among microbial or host taxa? What underlies any putative differences among taxa? Among all DNA transfer events in either direction, how many are retained and functional?

Interestingly, the effects of symbionts on the host genome extend beyond nuclear genes when the microorganism is co-inherited with non-nuclear DNA (Fig. 3). In particular, mitochondrial DNA (mtDNA) is co-inherited with some symbionts via the cytoplasm, and thus sometimes associates with specific bacterial or microsporidian endosymbiont infections that can lead to co-associations between symbionts and mtDNA haplotypes[162-164]. In these cases, certain mtDNA haplotypes become overrepresented in a population[165], mtDNA nonsynonymous mutations increase possibly in an arms race with the symbiont,[166] and mtDNA diversity is reduced compared with uninfected counterparts at either a population level[167] or globally[168]. Moreover, *Candidatus* Midichloria mitochondrii bacteria that infect ticks exhibit the unique ability to invade mitochondria of ovarian cells[169], although the exact purpose and effects are not known. Overall, the intertwined evolution of mtDNA and symbiont occur specifically when the symbiont is intracellular, and thus this association is common for the many different endosymbiotic bacteria. The association is much rarer for fungi that are less often intracellular[15], and it is unknown for

viruses and protists, which may or may not be a reflection of their biology and remains to be further explored.

The impact of host-microorganism symbiosis on microbial genome evolution can be substantial[14] (Fig. 3). Indeed, vertically transmitted and obligate intracellular bacteria frequently experience reductive genome evolution owing to the confined lifestyle[12], relaxed selection due to functional redundancy with hosts[170], genetic drift that occurs through the bottlenecks of vertical transmission through the matriline[171,172] and accelerated sequence evolution together with altered base compositions[173]. In general, in the cases of evolving mutualisms, genes may be lost in either the host or microorganism to avoid redundancy in the hologenome. The result is genome erosion until host and symbiont genomes complement each other[14], which can mean that one or more partners may be left without an essential gene, and the relationship becomes obligate. However, one standing question is how non-bacterial genomes change. For example, do intracellular, vertically inherited fungi or protists exhibit the same rates of gene loss and sequence evolution as bacterial symbionts? Are they more or less likely to experience genetic changes similar to bacteria, and why?

Host behavior, sexual selection and speciation

Given the dependency of many microorganisms on host reproduction to spread through a population, they can influence host mating behavior and mate choice to facilitate their transmission. For example, *Wolbachia* infection can lead to increased female promiscuity and male fatigue in sex-biased, male-killer populations[174], discrimination between infected and uninfected sister species to avoid cytoplasmic incompatibility lethality[175] as well as preference for mating between uninfected individuals[52]. One particularly striking case is the effect of male-killing *Wolbachia* in *Acraea encedon* butterfly hosts[176]. Populations can become extremely female biased because of high infection rates with the male killer. With fewer opportunities to mate, females begin to form lekking swarms and exhibit mate-attracting behaviors, a departure from the canonical formation of male lekking swarms to attract females. This inverted form of sexual selection ultimately enables males to be selective about preferable female characteristics, whereas normally the opposite is true[176].

Importantly, manipulation of host reproductive behavior to facilitate microbial spread goes beyond mate discrimination and mating frequencies. Indeed, infection of cicadas with the fungus

Massospora cicadina correlates with altered male wing-flick patterns that mimic females so other males are attracted and infected[17], and infection of *Helicoverpa zea* moths with the gonad-specific Hz-2V virus correlates with a fivefold to sevenfold increase in female sex pheromone production and increased mating calls to attract and infect males[177]. Particularly for microorganisms that depend on host reproductive tissues to transmit, altering mating behavior through direct manipulation or indirect mechanisms may be a successful survival strategy. However, many questions remain. Are microorganisms of reproductive tissues more likely to influence host mating behavior than symbionts of other body sites? Do they resultantly have a unique influence on host sexual selection that other symbionts less often have? Is the impact the same among parasitic and beneficial, exclusively and non-exclusively vertically inherited, or gonad-specific and multi-tissue symbionts? What are the mechanisms that drive behavioral changes; are they direct or indirect interactions? Are there differences among microorganisms of different classifications? What effects do these behavioral changes have on the rest of the microbial community over many host generations?

With the profound influence that symbionts may have on arthropod mating behaviors and reproduction, it follows that in some cases, they can contribute to host reproductive isolation and thus speciation[178]. Mechanisms of symbiont-induced isolation or speciation may include mate discrimination based on infection status[175], hybrid sterility from microbial over-proliferation[179,180], hybrid lethality[181,182] or reproductive isolation owing to microorganism-mediated specialization in distinct niches[134]. For example, closely related species of the parasitoid wasp genus *Nasonia* that diverged several hundred thousand to one million years ago are strongly reproductively isolated by cytoplasmic incompatibility-inducing *Wolbachia* that cause severe lethality of F1 hybrid offspring in interspecific crosses[181]. However, this general phenomenon is most-often demonstrated in the cases of bacterial symbionts, and it is less understood to what extent other microorganisms may play a part. In at least the cases involving *Wolbachia*-induced cytoplasmic incompatibility, a role of the phage WO genes has been established[158]. Many additional questions still need to be addressed: what is the frequency of microbial involvement in speciation events, do the mechanisms differ for different microorganisms, and is the contribution to host speciation enriched among parasitic or gonad-specific symbionts?

Conclusions and Outlook

Microorganisms of host reproductive tissues have unique relationships with their hosts. Their proximity to germline tissues enables a greater probability of interaction with hereditary DNA and vertical transmission that generates a dependency on host fitness. Moreover, the heritability of microorganisms in these tissues raises the potential to have multi-generational impacts that span from individual physiological effects to speciation. These characteristics have led to interactions with the host that are unique or may occur with a different frequency than among microorganisms of other tissues. Notably, the principles from arthropod–microorganism interactions in host reproductive tissues can extend to both humans (Box 2) and plants (Box 3). Building on substantial work within the field, many areas for future research are important, since a wide context is still missing. We understand much regarding the bipartite interactions among certain hereditary bacteria–host pairs, but far less is known about microorganisms that fall under different criteria. To better understand the fundamental biology of microorganisms in the unique context of the reproductive tissues, the field will benefit from vigorous attention to the greater diversity of microorganisms and hosts in their full ecological contexts.

Future research should emphasize investigation into the identities of non-bacterial microorganisms in the reproductive tissues of diverse host taxa, entire microbial communities in the reproductive tissues, and the transient microorganisms of the reproductive tissues. Much research interest has focused on microorganisms such as *Wolbachia* that have importance in vector control[183], but research into other diverse organisms often remains scarce. Non-bacterial microorganisms interact with hosts in ways that are both similar (such as reproductive parasitism[6]) and different (such as phage manipulating both bacterial and arthropod hosts[158], or more frequent paternal transmission among certain taxa[86]) from the ways that bacteria do. Therefore, investigation into the identity of all microorganisms, their population dynamics, mechanisms of interactions with host and other microorganisms, and their functional evolutionary consequences will be crucial in the future. This is especially important to do across diverse host taxa to provide a more comprehensive perspective and framework to identify major biological themes across nature, since model organisms or those of importance to agriculture and health have thus far received proportionally greater attention.

Regarding communities in reproductive tissues, it will be important to not only characterize their identities, but also any putative "core" microbiota that could be important for studies of evolution, vector control initiatives, or conservation efforts. Many microbiota subfields have flourished in recent years[56,184-186]; however, the arthropod reproductive tissue context is currently underexplored. Although whole body microbiota analyses on arthropods are not lacking in general[43,187,188], few studies have specifically characterized the microbiotas of reproductive tissues[9,44]. In addition, many studies lack proper, sequenced contamination controls or assays to assess if DNA is from living microorganism, so current findings must be taken with caution, and future studies should include such controls. Building on the research foundation of bipartite symbioses, it will be important for the field to interrogate complexity in microbial networks to gain a more holistic understanding of the microorganisms in reproductive tissues.

Finally, of the microorganisms that have been most fully explored in the literature, most are hereditary or pathogens such as endosymbionts[189] or the infectious agents of sexually transmitted infections[16,190], respectively. However, very little is known about other members of the arthropod reproductive tissue microbiota (including opportunistic microorganisms, reviewed elsewhere[90]), particularly those that are horizontally acquired or whose host phenotypes caused by infection are less pronounced. For example, it is important to discover how much of the microbiota is horizontally or vertically acquired, whether these frequencies differ among different microbial taxa or hosts, what factors determine the establishment or loss of symbionts in the host, how or whether newly acquired symbionts interact with hereditary microorganisms, or what level of selection there is for these transient organisms. It will also be important to continue research on microorganisms that transiently colonize the reproductive tissues versus those that exclusively colonize them. How are their rates of genetic exchange with the host different? Are there biases in which genes are exchanged? Do they have broadly similar impacts on host fitness and evolution? How are these symbioses different, in terms of both the relationship with the host and other microorganisms in the reproductive tissues and throughout the body? Are they more or less likely to exhibit parallel phylogenies with the host?

Continued research to answer these questions will enable advancements not only in our understanding of fundamental biological principles, but also potentiate new applied research in areas of vector biology, agriculture and conservation.

Box 2: Microorganisms of human reproductive tissues

Unlike arthropods, characterization of the microbiota of human reproductive tissues is common, but less is understood about their function or evolutionary consequences. In humans, bacteria are the best studied and are the most abundant in reproductive tissue samples[191]. The vaginal microbiota also contains a smaller proportion of diverse fungi[192] and viruses[193], but non-bacterial microorganisms are not fully characterized in men and women and merit further study. In contrast, bacteria are well-studied, particularly within women. Bacterial 16S rDNA gene sequencing has revealed that, within groups of reproductive-age women that have been studied thus far, the vaginal microbiota is diverse and often dominated by *Lactobacillus* species, including *Lactobacillus crispatus*, *Lactobacillus iners*, *Lactobacillus jensenii* and *Lactobacillus gasseri*[191]. In a minority of women tested in these studies, the vaginal microbiota comprises diverse anaerobic bacteria such as *Streptococcus* with no single dominant species, whereas others are dominated by anaerobic bacteria such as the *Prevotella*, *Atopobium* and *Gardnerella* genera[191]. Many factors contribute to inter-individual variation in the vaginal microbiota, including ethnicity[191], pregnancy[194], menopause[195], menstruation[196], hygiene[197], use of birth control[197] and age[197]. In addition, women contain microorganisms in breastmilk that are important for offspring health[198].

Less is known about the reproductive tract microbiota in men compared with women. However, the upper male genital tract (prostate and vas deferens) is considered to be germ free except during infection[199,200] and the lower genital tract (urethra and coronal sulcus) contains dozens of bacterial families, of which Clostridiales and Prevotellaceae are most abundant[201]. Common genera include *Corynebacterium*, *Anaerococcus*, *Staphylococcus* and *Prevotella*[201], which indicates some overlap between genera found in the female reproductive tracts, whereas some are more common in men. 16S rDNA gene sequencing indicates that there is likely no 'core' penis microbiota[202], a result that parallels the extensive inter-individual variation discussed above for the vaginal microbiota. In addition, circumcision[202], prostatitis[203] and prostate cancer[200] associate with changes in the reproductive tract microbiota.

Microorganisms in the reproductive tract also have an impact on human health and fitness. For example female genital tract infections (GTIs) are associated with pelvic inflammatory disease, ovarian abscesses, tissue scarring, and infertility among other conditions[204]. Other microbial infections can also cause complications in pregnancy and birth such as stillbirth, sepsis or preterm birth[205]. In addition, disruption of vaginal microbiota homeostasis is a hallmark of

bacterial vaginosis, whereby the community changes from *Lactobacillus* dominance to more diverse communities of mostly obligate anaerobes[206]. Therefore, *Lactobacilli* are likely to have important beneficial roles, but it is unclear what functions other microorganisms may have. Microorganisms in the male reproductive tract also affect male fitness. Indeed, the abundances of certain microbial genera such as *Prevotella* and *Pseudomonas* are associated with poor semen quality, including reduced motility, volume and concentration, whereas the abundances of *Gardnerella* and *Lactobacillus* species are associated with normal semen quality[207]. Additionally, GTIs in men correlate with poor sperm quality and are associated with infertility[204]. Bacterial DNA sequences are detected in semen of healthy men[203], so future studies could determine if any living, resident microorganisms confer benefits to male reproduction. In addition, long-term evolutionary consequences of these symbioses on the host are poorly understood compared with arthropod symbioses and remain open questions.

Box 3: Microorganisms of plant reproductive tissues

Plant reproductive tissues contain a diverse range of microorganisms. The bacterial microbiota of plant reproductive tissues includes various endophytes (microbial symbionts that inhabit plants without any apparent harm) and phytopathogens (microbial symbionts that parasitize plants). In addition, the plant microbiota comprises many known viral[208] and fungal[209] pathogens, as well as many protective fungal endophytes (reviewed in Ref. [210]). Among bacteria, the best surveyed group of plant microorganisms, a comprehensive compilation of seed endophytes from 2014 spans 131 bacterial genera and four phyla that occur in 25 different plants[211]. The most common phyla were Proteobacteria, followed by Actinobacteria and Firmicutes. Cultivation or sequencing methods across multiple studies demonstrated that 17 of the plant species seeds contained multiple phyla, whereas the remaining eight only contained a single bacterial phylum[211]. A plant seed may also contain several species within the same phylum or genus[212]. Bacteria commonly sequenced in seeds include various species of *Pantoea*, *Methylobacterium*, *Bacillus*, *Stapylococcus*, *Pseudomonas* and *Paenibacillus*[211,213-215]. Notably cultivation methods in many plant species result in only one or a small number of bacterial species[216]. This may either reflect the inability to culture most microorganisms of seeds, or the presence of few bacterial cells. Known factors affecting the sequence diversity of seed microbiota include plant species[211], location outside or inside the seed[217], and the stage of seed maturation[217]. Other reproductive structures of

plants have site-specific microbial communities as well, such as flowers[218], fruits[213] and pollen[219]. For example, pollen of diverse host species has many common bacterial genera, including *Rosenbergiella*, *Pseudomonas*, *Methylobacterium*, *Friedmanniella* and *Bacillus*[219-221], which represent some shared and distinct symbionts compared with those commonly found in seeds.

Interestingly, plant-microorganism symbioses are known to share many of the same evolutionary principles demonstrated in arthropod-microorganism symbioses. For example, symbiotic modulation of host gene expression in reproductive tissues is known in plants. One extraordinary case is that of phytoplasma, which are obligate intracellular bacteria that are transmitted from plant to plant via insect vectors, such as leafhoppers and psyllids[222]. Phytoplasma cause symptoms such as yellowing of leaves, greening of flowers or even sterility[222]. In *Arabidopsis thaliana*, phytoplasma produce an effector protein, SAP54, which changes the flowers into leaf-like vegetative structures[223]. Both SAP54 and phytoplasma effector protein SAP11 manipulate plant host transcription factors that regulate normal flower development[224]. Leafhoppers then preferentially choose infected plants for oviposition owing to the physical changes in the plant[223]. This enables the bacteria to spread by attracting its insect vector. In addition, bacterial seed endophytes in plants can also determine the environmental niche of the host. For example, endophytic bacteria of the giant cardon cactus help seedlings develop on sites of barren rock[225]. When the endophytes are eliminated by antibiotic treatment, the seeds fail to develop, but when sterile seeds are re-inoculated with the endophytes, their growth is rescued[225]. The endophytes are able to accomplish this remarkable feat by performing two tasks: fixing nitrogen for the host, and producing various organic acids that weather the rock and release the life-sustaining minerals required for growth in this harsh environment[226]. In addition, vertically inherited bacterial seed endophytes of the rattlebox shrub include microorganisms that contribute to growth and resilience, which probably promotes the ability of the host to survive in metal-contaminated mining sites[227]. Furthermore, some of these beneficial relationships may result in microbial genome erosion and development of obligate associations, as described for arthropods. One example is the case of *Burkholderia* endosymbionts of *Psychotria* plants that are vertically transmitted and are likely to have a role in protection from predation or pathogens[228]. Additionally, it is known that microorganisms of different taxa may cooperatively interact in host reproductive tissues. For example, members of the fungal genus *Rhizopus* that cause rice seedling blight contain their endosymbiotic bacteria that produce a phytotoxin that is critical in pathogenesis. When the

toxin damages host tissues, both the fungal host as well as the bacterial symbionts of the fungi benefit from plant nutrients, in a striking case of inter-taxa dependence and function in parasitization of host reproductive tissues[229]. Thus, plant-microorganism symbioses in reproductive tissues of plants share many important evolutionary characteristics to that of arthropod-microorganism symbioses and merit further study.

Acknowledgements

The authors thank M. I. Hood-Pishchany, K. Ngo, B. Leigh and the journal editorial team for constructive comments. Work in the authors' laboratory was supported by awards from the National Institutes of Health (NIH; R21 AI133522) and the Vanderbilt Microbiome Initiative to S.R.B, as well as by NIH grant F31 AI143152 to J.I.P. *Nature Reviews Microbiology* thanks G. Hurst, I. Newton and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Author Contributions

S.R.B. and J.I.P. wrote the article, reviewed and edited the manuscript before submission. J.I.P initially drafted the article and researched data for the article.

Microbe-mediated male killing

Contributing Authors: Jessamyn I. Perlmutter, Jane E. Meyers, and Seth R. Bordenstein

Abstract

The most widespread symbionts on the planet include a variety of microorganisms that manipulate host reproduction. Among these are a diverse group of bacteria, viruses, and fungi that hijack arthropod host reproductive processes to facilitate their own spread. There are a variety of methods they employ to control the host, including several that manipulate host sex ratios to increase the fitness of the females that transmit them. Male killing is one such sex ratio-distorting manipulation where sons of infected females are selectively killed. Published research on male-killers spans a century, and in that time, many discoveries have advanced our understanding of male killing. Here, we synthesize a century of work on the male killing phenotype, focusing on its diversity, evolutionary impacts on host and microbe, and the genes and mechanisms underlying the phenotype. In addition, we highlight remaining questions and important areas for future research and discuss potential applications of male killing in arthropod control and conservation.

Introdution

One century ago in 1920, the self-trained Pacific entomologist, Hubert W. Simmonds, accepted an appointment as an official agricultural entomologist for the government of Fiji[230,231]. His interest in insects began in childhood in England and grew into a lifelong passion that built his career while living on many islands throughout the South Pacific. There, he observed the diversity of arthropods in nature and created large collections of insects, published experiments, and led efforts to introduce predatory insects to help control house flies and various agricultural pests [230,231]. During his decades in Fiji, Simmonds made observations on the characteristics of many local insects, including the captivating wing pattern polymorphisms of the blue moon butterfly (*Hypolimnas bolina*), among other *Hypolimnas* butterflies (Fig. 2-1)[231]. Here, he made one of the first discoveries of a female-biased population of butterflies that would be the subject of his and other studies for years to come[232-234]. Indeed, many of these studies predated the initial discovery of *Wolbachia* bacterial genus now known to cause reproductive phenotypes such as male killing[235]. Indeed, Simmonds discovered through butterfly breeding that not only was male killing heritable,

but also that half of the eggs laid died, presumably the males. This led to the hypothesis of a heritable all-female, embryonic, son-killing trait from the initial observation of a sex bias in the population[234]. This was not the only discovery of sex-biased insect populations at the time, however. In 1914, entomologists E. B. Poulton and W. A. Lamborn similarly reported a relative rarity of male *Acraea encedon* butterflies in various populations in Africa[236]. Several families that were bred in captivity exhibited an all-female, son-killing trait just as the *Hypolimnas* butterflies halfway across the world in Fiji had[236].

Figure 1-4. Hand-drawn wing pattern of *Hypolimnas inoptina* by Hubert Simmonds.
From his autobiography, "My Weapons Had Wings"[231].

What began as simple observations by a hobbyist turned career entomologist and others from the early 20th century then grew into a field of scientific research on microbe-host interactions that continued for more than 100 years. Over the ensuing decades, entomologists conducting scientific research would report intriguing heritable sex ratio biases in various insect populations. It began with early empirical works including those by Lusis on *Adalia bipunctata* ladybugs in 1947[237], Magni on *Drosophila bifasciata* flies in 1953[238], Cavalcanti and Falcão on *Drosophila prosaltans* flies in 1954[239], Carson on *Drosophila borealis* flies in 1956[240], and Malogolowkin and Poulson on *Drosophila willistoni* flies in 1957[241]. At this point, it was recognized that such sex ratio biases were cytoplasmically-inherited and likely of infectious origin. Indeed, Malogolowkin and Poulson performed experiments wherein the ooplasm of infected eggs was extracted and injected into the abdomens of adult females. From this, five of 16 injected females began to produce sex-biased offspring, and their female offspring also inherited the trait, supporting the

conclusion of the infectious nature of the sex ratio bias[241]. Further, Poulson and Sakaguchi reported in 1961 that sex ratio bias in *D. willistoni* and *Drosophila nebulosa* was associated with spirochetes in the hemolymph of adult females[242], and later showed that this trait (and the spirochetes) could be transferred to *Drosophila melanogaster* via microinjection[243]. A few years later, in 1970, it was first proposed by Kugao Oishi that the spirochetes produced a fabled "androcidin" toxin that was responsible for male death in *Drosophila* hosts[244]. Inherited sex ratio biases were commonly found, and an understanding of their infectious origin began in the mid-century and continued to mature over the decades. By the end of the century and with the advent of genetic surveys, *Wolbachia* bacteria were reported as a male-killer of *Adalia bipunctata* ladybugs and *Acraea encedon* butterflies[245], and *Spiroplasma poulsonii* was named as the spirochete underlying many of the sex ratio biases identified in neotropical fruit flies[246]. It was not until 2002, over 80 years since the initial observations by Simmonds, that *Wolbachia* was named as the male-killer of *Hypolimnas bolina*[247]. By 2018, the existence of a *Spiroplasma* toxin hypothesized by Oishi was confirmed with the discovery of the *Spiroplasma poulsonii* SpAID (*Spiroplasma* androcidin) toxin[248,249], and a gene candidate for *Wolbachia*-induced male killing, *wmk* (*WO-mediated killing*), has also been identified[160].

Over the past century, diverse male killing agents and hosts have been classified, and much progress has been made in our understanding of their ecology, evolution, and molecular mechanisms. Indeed, "reproductive manipulators" are now known to cause many different phenotypes including feminization (genetic males develop and reproduce as females), parthenogenesis (asexual reproduction of females), and cytoplasmic incompatibility (CI, offspring die when infected fathers are crossed to uninfected mothers, but live when mothers are also infected)[23]. These agents are often maternally-inherited and help increase the fitness of infected females. Male killing is another such manipulation, and terms such as "sex ratio agents", "son-killers", or "male-killers", collectively refer to the increasingly diverse group of parasitic microorganisms that selectively kill male hosts. As technologies advance, new symbioses are discovered, and interest rises in new and improved pest and vector control strategies, so does our knowledge of male killing. Here, we summarize and discuss the current status of the field after a century of research. We cover (i) the diversity of male-killing symbioses, (ii) their impact on the evolution of both host and microbe, (iii) progress on the genetics and mechanisms of male killing,

and (iv) potential applications of the phenotype. We also highlight research questions in need of further attention and discuss important future directions in the field.

Diversity of Male Killing

Although selective death of male hosts in sexually-reproducing populations seems a bizarre or counterintuitive approach to survival, it is one of the most widely-evolved reproductive manipulations by microbes. Indeed, an array of bacteria (*Wolbachia*, *Rickettsia*, *Arsenophonus*, *Spiroplasma*, unnamed Flavobacteria)[23], viruses (unnamed RNA viruses and potentially prophage)[28,160], and fungi (microsporidia including *Thelohania* and *Ambylospora spp*.)[250,251] can kill male hosts. Further, the hosts themselves represent many major insect orders, including Coleoptera (beetles[252]), Diptera (flies[35,250]), Lepidoptera (butterflies[253] and moths[125]), Neuroptera (lacewings[254]), Hymenoptera (ants[255], wasps[256]), as well as arachnids (mites[257] and pseudoscorpions[26]), and are found on every continent except Antarctica (Table 2-1). Based on the variety reported thus far, researchers have speculated that male killing may be an easy trait to evolve[130,258]. Indeed, this list is unlikely to be representative of the whole diversity of male-killers nor their hosts as there has been no comprehensive search for the phenotype in nature, and studies thus far are biased due to geography, importance of the host in pest or vector control, ease of rearing in the laboratory, ability to visually distinguish host males and females, and strength of the phenotype (ie, the proportion of males that die, and how common the male-killer is in the population). Even so, a great diversity has been discovered. It is thought that male killing is easy to evolve not only due to the number of host-microbe symbioses that exhibit male killing, but also due to their diversity in terms of host phylogeny, microbial taxonomy, geographical location, phenotypic traits, host sex determination systems, and microbial male-killing genes and candidates identified thus far (Table 2-1, or described below). Indeed, it is likely that male killing has independently evolved in different systems many times due to the extensive variation in affected host-symbiont pairs. Future work should focus on identifying male killers in additional hosts in previously unexplored or underexplored groups. In particular, are there other fungi that kill males (so far, only microsporidia in mosquitoes are recognized) or other viruses (so far, only a few identified at all, and they are relatively uncharacterized)? Are there other major insect orders or other kinds of arthropods (additional arachnids, crustaceans, etc.) that have male-killing symbionts? Are they more common to hosts with certain characteristics like shorter lifecycles or

those that undergo diapause than others (it is already proposed to be more common in insects that lay large clutches of eggs[259])? Do certain climates promote maintenance of male killing over others? How does the prevalence or incidence of male killing compare to other reproductive manipulations in nature? Is male killing more taxonomically widespread than other manipulations, and why? Indeed, more symbiont taxa are known that can induce male killing than other manipulations such as CI, caused by *Wolbachia*[260] or *Cardinium*[27], which means that in terms of taxon diversity, male killing could be more common than any other reproductive phenotype.

Table 1-1. Reported cases of microbial male killing.
This list is not comprehensive in any one category, but serves as a reference for most known cases of male killing by microorganisms. In addition, some cases are suspected but unconfirmed male-killers, and are included due to having a reported biased sex ratio.

Bacterial MK				
	Symbiont	**Host Species**	**Location**	**Citation**
	Wolbachia pipientis	*Acraea encedana*	Sub-Saharan Africa	[167]
		Acrea encedon	Sub-Saharan Africa	[176]
		Adalia bipunctata	Russia, Sweden	[245]
		Coccinella undecimpunctata	Middle East	[261]
		Cordylochernes scorpioides	Panama	[26]
		Drosophila bifasciata	Japan	[262]
		Drosophila borealis	USA	[263]
		Drosophila innubila	USA, Mexico	[264]
		Drosophila pandora	Australia	[265]
		Drosophila subquinaria	USA	[266]
		Ephestia kuehniella	Japan	[267]
		Homona magnanima	Japan	[268]
		Hypolimnas bolina	Pacific islands	[247]
		Ostrinia furnacalis	Japan	[269]
		Ostrinia orientalis	Japan	[270]
		Ostrinia scapulalis	Japan	[271]
		Spodoptera exempta	Sub-Saharan Africa	[272]
		Tribolium madens	Europe, Canada	[252]
	Rickettsia sp.	*Adalia bipunctata*	Northern Europe	[273]
		Adalia decempuctata	Europe	[274]
		Adalia 10-punctata	Northern Europe	[274]
		Bracys tesselatus	USA	[275]
		Propylea japonica	Japan	[276]

	Arsenophonus nasoniae	*Nasonia vitripennis*	USA	[256]
	Spiroplasma ixodetis relative	*Adalia bipunctata*	Northeastern Europe	[277]
		Acyrthosiphon pisum	UK	[278]
		Anisosticta 19-punctata	England	[279]
		Cheilomenes 6-maculata	Japan	[280]
		Danaus chrysippus	Asia, Australasia, Africa	[281]
		Harmonia axyridis	Northern Asia	[282]
		Homona magnanima	Japan	[283]
		Laodelphax striatellus	Japan, Taiwan	[284]
		Menochilius sexmaculatus	Australasia	[285]
		Ostrinia zaguliaevi	Russia, Japan, China	[286]
	Spiroplasma poulsonii	*Drosophila nebulosa*	Brazil	[287]
		Drosophila neocardini	Brazil	[288]
		Drosophila melanogaster	Brazil	[289]
		Drosophila ornatifrons	Brazil	[288]
		Drosophila paraguayensis	Brazil	[288]
		Drosophila willistoni	Americas	[246]
	Other or Unknown *Spiroplasma spp.*	*Mallada desjardinsi*	Japan	[254]
		Drosophila equinoxalis	Central and South America	[290]
		Drosophila paulistorum	South America	[291]
		Drosophila prosaltans	Central and South America	[292]
		Drosophila robusta	USA	[293]
	Flavobacteria	*Adonia variegata*	Europe	[294]
		Coccinula crotchi	Japan	[261]
		Coccinula sinensis	Japan	[295]
		Coleomegilla maculata	USA	[296]
		Harmonia 4-punctata	France	[261]
	Other or unknown bacteria	*Cadra cautella*	Japan	[297]
		Coccinella 7-punctata	England	[261]
		Coccinella 7-punctata brucki	Japan	[261]
		Calvia 14-punctata	England, Canada	[298]
		Gastrolina depressa	Japan	[299]
		Lymantria dispar	Japan	[300]
		Spilostethus hospes	Australia	[301]
	Orientia tsutsugamushi	*Leptotrombidium arenicola*	Malaysia	[257]
	Orientia tsutsugamushi	*Leptotrombidium fletcheri*	Malaysia	[257]
Microsporidial MK	*Ambylospora* spp.	*Aedes, Culex, Culiseta spp.*	Worldwide, tropical	[250,302,303]
	Parathelohania legeri	*Anopheles quadimaculatus*	USA	[304]

	Parathelohania obesa	*Anopheles quadimaculatus*	USA	[305]
	Thelohania spp.	*Aedes, Culex, Anopheles spp.*	Worldwide, tropical	[304,306]
	Parathelohania anophelis	*Anopheles quadimaculatus*	USA	[307]
Viral MK	Unknown RNA virus	*Drosophila biauraria*	Japan	[28]
		Homona magnanima	Japan, Southeast Asia	[125]
Unknown or Unconfirmed Agent		*Attagenus anicolor japonica*	Japan	[308]
		Calligrapha philadelphica	USA	[309]
		Caraphractus cinctus	Europe	[310]
		Dendroctonus jefferyi	USA	[311]
		Epiphyas postvittana	Australasia	[312]
		Estigmene acrea	Africa, Americas	[313]
		Hippodamia quinquesignata	USA	[314]
		Ips latidens	USA	[315]
		Lophyrus pini	Europe	[316]
		Oncopeltus fasciatus	North and Central America	[317]
		Oncometopia nigricans	North America	[318]
		Orthotomicus latidens	North America	[315]
		Phyllonorycter sorbicola	Japan, Russia	[319]
		Pieris napi	Europe	[320]
		Pithyophtorous sp.	USA	[315]
		Pygaera pigra	Europe, North Africa	[321]
		Spodoptera littoralis	Europe, Africa, Middle East	[322]
		Tetranychus urticae	Worldwide	[323]
		Xyleborus sp.	Worldwide	[324]

There are many factors that influence the discovery of male killing in a host population. Importantly, there is a wide variety of trait penetrance, with some cases exhibiting complete transmission, death of all sons, or high population prevalence, and others exhibiting lower transmission, some surviving sons, or low population prevalence[258,261,264]. Indeed, many populations have few individuals infected with the male-killer, so cases may be missed in population surveys unless a host with the male-killing symbiont is isolated and reared in the lab. For example, different populations of *D. bifasciata* in Japan may only have infections at a rate of 0-7%[325]. Further, it may be more overlooked in hosts with certain lifecycles that are less often researched than others, and thus many cases may be overlooked. For example, the only known

case of male killing in a live-bearing species is by *Wolbachia* in the harlequin beetle-riding pseudoscorpion, wherein male embryos developing in the mother's brood sac selectively die[26]. In addition, few hemimetabolous species are known to harbor male killers, with *A. pisum* aphids infected with a *Spiroplasma ixodetus* relative being one of the few[278]. These factors, among others including those described above, greatly influence our knowledge of the diversity of male killing. Future work should continue to identify populations with low male-killer prevalence and those with underrepresented lifecycle traits or reproductive strategies.

Beyond diversity in the taxa of male-killers and hosts, there is diversity in the number and taxa of male-killers within a host or population. Indeed, several arthropod populations are infected with multiple strains of the same taxon, or several different taxa. For example, *Homona magnanima* moth populations can be infected with male-killing *Wolbachia*[268], a *Spiroplasma* male-killer[283], or an RNA virus male-killer[125], and in some cases they are infected with multiple endosymbionts in the same individual that may or may not include a male-killer[283,326]. In addition, some species such as *Adalia bipunctata* ladybugs can be infected with many different male-killers, with single populations harboring *Rickettsia*, *Spiroplasm*a, or two different strains of *Wolbachia*[273]. Other species are only known to be infected by a single microbial taxon, and there may be variation in the susceptibility of a particular host species to male killing by a particular strain. For example, in the *D. recens* and *D. subquinaria* sister species, *D. recens* harbors a *Wolbachia* strain that causes cytoplasmic incompatibility, but when it is transferred to the naturally-uninfected *D. subquinaria* via introgression, it induces male killing. In addition, this trait is polymorphic, as some strains of *D. subquinaria* are susceptible to male killing, while others are not[266]. Although these studies have revealed the variety of male-killing infections that may affect an individual or host population, relatively little is known about the dynamics among these symbionts. For example, it is unknown if multiple male-killers infect the same individual (they co-occur with non-male-killing strains in the literature), and if so, how they interact. In addition, it is unknown how a male-killer and another reproductive manipulator may co-exist in the same individual, or how they may interact at a population level. Competition dynamics are likely among symbionts in an individual or population, but they remain underexplored. In particular, in populations with multiple male-killers, how do they interact? Are they cooperative or competitive? Future work will be needed to assess these dynamics and uncover the extent of the diversity of male killing in nature.

Evolution of Host and Male-Killing Microbes

Male killing is a curious strategy for survival in sexually reproducing hosts. Although some advantages of male killing are known, it is still difficult to explain how male killing is maintained in a population over time. If a male-killer is prevalent enough in a population and has efficient transmission, extinction of host and microbe or resistance and loss of the symbiont would be predicted[327]. Indeed, extinction is thought to have likely occurred in nature[258,327] and resistance to male killing is common among diverse hosts including flies[266], ladybugs[280], and butterflies[328,329] (Table 2-2). However, at least in some populations, male killing phenotypes or their symbionts have been maintained over long periods of time[264,330]. For example, the *Hypolimnas bolina* butterflies first reported to have a sex ratio bias by Simmonds in 1923[232] still exhibit a sex ratio bias to this day due to *Wolbachia* infection[331]. Indeed, the populations have shown fluctuating sex ratios for decades, indicating that although there may be changing infection dynamics across a few generations, the symbiont and host still remain. Some factors thought to influence maintenance over time include the theoretical advantage of male killing including prevention of inbreeding[332,333], allocation of resources to females where females consume dead siblings etc. (likely to occur in cannibalistic ladybugs[334], and hypothesized in pseudoscorpions where the mother's resources will be allocated to females in the sac when males die[26]), reduced competition for nearby food resources among females[335,336], imperfect transmission[264,327], or benefits such as protection from parasitoids[35] or host viruses[336]. Further, none of these advantages applies to every system. In some cases, the symbiont provides no protection against viruses[337], and in others there is no evidence for reduced inbreeding[336], and still others exhibit perfect transmission[167]. It has been previously postulated that each species may be more or less able to maintain a male-killing infection[263]. For example, *D. borealis* infected with *w*Bor *Wolbachia* around the Great Lakes in the United States and *D. innubila* infected with *w*Inn *Wolbachia* in the mountainous sky islands in the deserts of the Southwest United States have very closely-related symbiont strains (phylogenetically closest to each other), while the hosts diverged 25-42 mya[263]. However, there are no known populations that share the infection between them and *D. innubila* and *D. borealis* are not known to interact ecologically. The facts that the hosts are geographically and phylogenetically distant with a similar symbiont suggest that there was likely a spread and subsequent loss of the symbiont in the physical space between the populations, and they are all

that remain. Further, the *w*Inn population has been infected for a long evolutionary time[264]. The maintenance in some species or locations but not others suggests that some populations are more or less likely to maintain male-killers for reasons that are still not completely clear. It likely has to do with the factors already discussed, or other factors that affect transmission such as temperature[138,262,338-340], but modeling and empirical work so far demonstrate differences among populations that do not provide a consistent answer that can be used to predict population outcomes. Other factors that could explain maintenance or loss of male-killers over time include proclivity for resistance to male killing, climate, probability of horizontal transmission, or proportion of individuals with aberrant sex (intersex, etc.)[130]. Also, very few populations have been monitored over a period of decades, so more long-term studies will be necessary to fully assess the delicate ecological balance required for a stable male-killer.

Table 1-2. Cases of likely or known host resistance to male killing.
Reported here are known cases where heritable resistance has been reported. The resistance is either known from studies demonstrating heritable resistance or is inferred as a possibility due to absence of male killing during phenotype switching between hosts.

Symbiont	Host Species	Based On Phenotype Switching	Citation
Wolbachia	*Cadra cautella*	Yes	341
	Drosophila melanogaster	Yes	342
	Drosophila prosaltans	No	343
	Drosophila recens	Yes	266
	Drosophila simulans	Yes	342
	Drosophila subquinaria	No	266
	Ephestia kuehniella	Yes	344
	Hypolimnas bolina	No	331,345
	Ostrinia scapulalis	Yes	267
***Spiroplasma ixodetus* relative**	*Cheilomenes sexmaculata*	No	280

The question of how a male-killer evolves over long periods of time is not fully resolved. One possibility is extinction if the host develops resistance or if the male killing is pervasive in a population and not enough males survive. Although the *Wolbachia* infection in some male-killers is thought to be ancient (15,000-700,000 years old) according to estimates based on genetic diversity and rates of evolution, it is not known if the male-killing phenotype itself has been present for that entire time. Long-term male killing is only evident in cases of the few *Wolbachia*-infected populations that have been monitored over time spanning a range from a few decades to a century

(*H. bolina, D. borealis, D. innubila*). Therefore, it is not empirically known how long a male-killer can last in a population. However, modeling suggests a male-killer could theoretically develop into a mutualist in the case of evolution toward haplodiploidy in the host[346]. Importantly, transitions to another symbiotic state have not been reported and would represent a significant advance in our understanding of male-killing dynamics. In addition, differences among taxa are not well understood. In particular, it would be intriguing to study the male-killing viruses and fungi, as they have different strategies in the host. Namely, bacteria typically induce "early" male killing, where males are killed as embryos[347]. However, viruses and fungi often cause "late" male killing, where males are killed as larvae or pupae[125,251]. This is thought to be due to the mixed modes of transmission in these taxa, where viruses and fungi multiply into later stages of development to enhance horizontal transmission when the host dies[130]. While females will vertically transfer the symbionts to offspring, males die during later development stages and spread the infection horizontally. This difference in transmission compared to bacteria may affect the possible long-term outcomes of male killing in these systems. Even within bacteria, the methods of transmission may be different. *Spiroplasma* in *D. melanogaster* hijack yolk uptake machinery from their primary location in the hemolymph[61], and *Arsenophonus* in *N. vitripennis* aggregate to the ovipositor of an adult female and are deposited onto the surface of parasitized fly pupae so that offspring become infected after eclosion[77]. Therefore, it will be an important future direction not only to observe, predict, and empirically test outcomes of male-killing in general, but also to compare the evolutionary outcomes among various microbial and host taxa.

Crucially, the evolution of hosts is also affected by male-killers in many ways. Indeed, symbionts that interact with insect sex-determining genes are hypothesized to influence the evolution of host sex determination. When a microbe manipulates host reproduction, it is often at the expense of the host and this generates genetic conflict between host and microbe. The host and microbe would theoretically enter into an arms race with the host evolving resistance, potentially altering its sex-determining molecules, and thus leading to larger-scale changes in host sex determination over evolutionary time[348,349]. As noted earlier, there are many known cases of resistance to male killing determined by host background (Table 2-2). Intriguingly, some have suggested that symbionts that manipulate host reproduction may, in part, have driven the diversity of sex-determining systems among arthropods. Indeed, populations of haplodiploid *Asobara* wasps infected with parthenogenetic strains of *Wolbachia* can no longer sexually reproduce and rely on

the symbiont for production of offspring[126]. This is due to erosion of genes involved in sexual reproduction that occurred because of relaxed selection with long-term asexual reproduction under parthenogenesis[128]. In addition, modeling suggests that long-term male killing may lead to evolution of haplodiploidy[346]. In nature, there is indeed a large variety of sex determination systems among affected male-killing hosts, providing some correlation between the phenotype and diverse sex determination, so their evolution could potentially have a connection. Male killers are known in species that have XY (males heterogametic, such as *Drosophila*[289]), ZW (females heterogametic, such as *Hypolimnas*[328]), haplodiploid (males haploid and females diploid, such as *Nasonia*[347]), or XO (females have two sex chromosomes, males have one, such as most arachnids[26]) sex determination systems.

This diversity in male-killing host sex determination systems is further intriguing since the mechanism would theoretically rely on the symbiont producing a toxin that targets a critical, male-specific process. Even within one sex determination systems, there may be variation in expression of sex-determining genes. For example, in XY flies similar to *D. melanogaster*, the master sex regulator in the Sophophora group, *sex-lethal* (*sxl*) determines female development when expressed[350], but it is also expressed in males of the Virilis *Drosophila* group and thus plays a different role in these species[351]. In addition, other components of the sex determination pathway such as dosage compensation gene *maleless* (*mle*) in *Drosophila* are not involved in dosage compensation in all insects[352]. Despite conservation of some molecules in sex determination across many insects like the *transformer* (*tra*) gene[353], there is not enough similarity across all affected groups for it to be likely that there is a shared target. This lack of conservation in molecules and mechanism is intriguing and one of many reasons it is considered unlikely that there is a universal male-killer gene or mechanism. It appears that male killing is easy to evolve and that diverse cases are the result of convergent evolution. This leads to important future questions regarding the interactions between male-killers and host sex determination systems, the diversity in mechanism of killing, and the diversity of microbial and host genes involved in male killing.

Beyond changes in host sex determination or development of host nuclear-encoded resistance to male killing, hosts may experience other evolutionary genetic changes ranging from potential host extinction to changes in mitochondrial DNA (mtDNA) diversity. When a male-killer spreads in a population, it will coincidentally associate with the mtDNA haplotype(s) of whichever hosts it infects. In addition, a male-killer with complete transmission efficiency to daughters will

stop the flow of mtDNA from the infected population to the uninfected population, and thus may act as a barrier to gene flow between the two. For example, *A. encedon* and *A. encedana* butterfly populations infected with *Wolbachia* exhibit low mtDNA diversity, while uninfected individuals exhibit greater diversity and show more geographical structure[167]. Indeed, selective sweeps reducing mtDNA haplotype diversity in a population have been reported in *Spodoptera exempta* armyworms[272] and *Hypolimnas* butterflies[165]. With reduced mtDNA diversity, it remains to be explored how this affects the long-term health of an infected population and if lowered population fitness from lack of haplotype diversity may balance the infection in a population. It is also unclear how this may affect long-term evolution between the host nuclear, mitochondrial, and symbiont DNA, and it is unclear how such mtDNA sweeps may affect symbioses with other symbionts. In addition, these sweeps are known in *Wolbachia*-infected populations but remain unknown for other symbionts, so this is an important future direction.

In addition, hosts may exhibit significant changes relating to sexual behavior. Experiments predict that in populations with rare males, sexually-transmitted infections (STIs) will accelerate their spread in a population[354]. Researchers used *Spiroplasma* male-killers in *Adalia bipunctata* ladybugs and compared mating behaviors between infected and uninfected hosts to monitor the spread of a sexually-transmitted mite. Empirical and modeling results demonstrate that males and females are likely to mate more frequently in the presence of a male-killer that reduces the male:female sex ratio. In turn, this is predicted to lead to more efficient spread of STIs overall, even accounting for the observation that transmission from males reduces over successive matings[354]. In addition, in populations of *Acraea encedon* butterflies, *Wolbachia* can reach a prevalence of 95% or more. In these cases, there is a sex role reversal where females form lekking swarms and exhibit mate-attracting behaviors, while males are picky about preferable female characteristics, thus altering sexual selection in the population[176]. In addition, males may prefer uninfected to infected females, however, theoretical modeling predicts that the infection will remain stable (which it is in nature[328]) as long as males make mistakes in discriminating between infected and uninfected mates[355]. Interestingly, in *H. bolina* populations, *Wolbachia* infection also associates with increased female mating frequency, however this trend peaks at mid-level infection frequency of about 50-75%. At higher levels, male mating rate is high enough to induce fatigue and reduce the size of spermatophores, such that females increase their mating rate to receive more[174]. Thus, behavior and reproductive ecology can be profoundly altered by a male-killer.

However, this remains limited to *Wolbachia* systems, and mostly in butterflies. Future work will be needed to identify how mating behavior is affected in other systems with diverse hosts, microbes, and STIs. Indeed, evidence in *D. innubila* suggests that mate discrimination does not occur at least in this system, so there are likely some differences.

Male-Killing Genetics and Mechanisms

There are multiple lines of evidence that male killing has evolved more than once. Beyond those previously described (including the variety of sex determination systems), there are differences in the phenotypes of different male-killing systems. For example, the timing may differ from one case to another. Bacterial male-killers and those that infect species of *Drosophila* tend to induce "early" male killing during embryogenesis, often within just a few hours[289,356]. For example, male *D. bifasciata* embryos infected with *Wolbachia* exhibit profound defects including chromatin bridging and spindle abnormalities within just a few hours of being laid[356]. There are some cases where this is not true, such as *Wolbachia*-infected *Ostrinia* moths that die in late embryonic to early larval development; thus, there is some variability in the timing from one symbiosis to another[269]. In contrast, microsporidian and viral male-killers often cause "late" male killing, usually in late larval or pupal stages where the individual has reached roughly their maximum weight and size[251]. This is thought to be a method of maximizing the number of particles or spores so that when the males die, horizontal transfer of the infection is maximized. However, there may be a density dependence to timing of male killing. For example, there is a known density-dependence for male-killing in *D. bifasciata* flies where females that are heat or antibiotic-treated pass on fewer bacterial cells to their offspring, thus weakening or losing the phenotype in the next generation[262,338]. In addition, the timing may depend on host background, as *w*Rec *Wolbachia* of *D. recens* flies cause larval male-killing in the *D. subquinaria* sister species[266]. Despite this nuance, the different timing and evolutionary strategies by different microbial taxa continues to support the likelihood of multiple mechanisms.

In addition, there are phenotypic differences in the embryos of different male-killers that suggest different mechanisms. For example, *Spiroplasma poulsonii* male-killers in *D. melanogaster* are known to cause several defects that do not occur in *Wolbachia* male-killers. *Spiroplasma*-infected fly embryos exhibit unique neurological defects, and *Wolbachia*-infected *D. bifasciata* embryos exhibit normal neural development[357]. In addition, while dosage compensation

markers are mislocalized beyond the X chromosome in male embryos with *Spiroplasma*[358], they remain concentrated on the X chromosome in *Wolbachia*-infected embryos[357]. However, these markers overlap with DNA damage in *Wolbachia*[357], suggesting that although the two infections may differentially affect male dosage compensation, they share a common tie to it. Indeed, there is substantial evidence that *Spiroplasma* male killing functions through disruption of the male-specific dosage compensation complex (DCC, also called Male Specific Lethal or MSL complex)[358-360]. In males deficient for various DCC proteins (which normally survive to late larval stages), *Spiroplasma* is no longer able to induce embryonic male death[359]. Coupled with the ability of *Spiroplasma* to kill females that artificially express the DCC[358], the evidence suggests that *Spiroplasma* requires the DCC to kill males. There is less evidence in *Wolbachia*, but the overlap of DNA damage and dosage compensation markers[357] and the fact that DNA defects accelerate when dosage compensation begins in male embryos (~3 h old)[356] both suggest *Wolbachia* also targets dosage compensation, although perhaps in different ways. In addition, both *Wolbachia*- and *Spiroplasma*-infected male embryos exhibit increased apoptosis during early development[357,360,361], although *Spiroplasma* reportedly induces greater levels of apoptosis earlier in development and the neural defects are independent of apoptosis[357]. In addition, both infections lead to DNA damage specifically on the male's dosage-compensated X chromosome seen in the form of a chromatin bridge[357,360]. The shared involvement of dosage compensation, early embryonic defects, chromatin bridging, and induction of apoptotic pathways coupled with the differences in degree of apoptosis and neural defects suggest that though they may share a common target, there are likely differences in the exact mechanism. This may be through targeting of different components of dosage compensation, among other possibilities, however this remains to be empirically tested.

The findings above in dipterans contrass with the work from lepidopterans infected with *Wolbachia*. Research on *Ostrinia scapulalis* moths demonstrates a substantially different method of male death. This species has a ZW sex determination system, rather than the XY system in flies, and only the ZZ males die[269,362,363]. Interestingly, when treated with antibiotics, both intersex and single sex offspring (male and female) were produced[362]. Although physically intersex based on coloring, these individuals had ZZ chromosomes (male) and had both male- and female-specific splice variants of the *doublesex* (*dsx*) gene involved in sex determination[364]. Later work determined that *Wolbachia* kills males through lethal feminization (creating female *dsx* splice variants), but at

lower levels due to antibiotics, males either develop normally or with some female characteristics that are not quite at lethal levels for their genotype[365]. Further intriguing is that when *Wolbachia* are completely eliminated, only males appear, but some are genetically female. This case represents an extraordinary symbiosis where *Wolbachia* compensate for loss of a critical feminizing factor in host females, but also kill males through feminization in a rare potential case of mutualism from a male-killer[269,363,365]. Additional research on *Ostrinia* moths shows that *Wolbachia* interfere with dosage compensation and associate with downregulation of a critical masculinizing gene (*Masc*)[366,367]. Beyond the inherent intrigue, this moth-*Wolbachia* symbiosis also stands in contrast with the mechanisms of male-killing in flies. Although all ultimately kill male offspring and both have connections to dosage compensation, the symbiont of the lepidopteran host does so uniquely through feminization and downregulation of masculinization, whereas there is no such evidence in drosophilid male-killers. This represents another likely case of convergent male killing through unique mechanisms.

In addition to contrasting findings of phenotypes, recent genetic breakthroughs have demonstrated the likely unique origins of *Spiroplasma* and *Wolbachia* male killing. The *Spiroplasma androcidin* (*Sp*AID) toxin was recently identified as the male-killing toxin of its symbiont in *D. melanogaster*[248,249]. When transgenically expressed, *Sp*AID recapitulates apoptosis, neural defects, and male-specific death. In addition, the *WO-mediated killing* (*wmk*) gene was recently identified as the strongest candidate for *Wolbachia* male killing in *D. melanogaster*[160]. When transgenically expressed, it too recapitulated many aspects of *Wolbachia* male killing including induction of a female-biased sex ratio, early embryonic defects like natural infections, and DNA damage that physically overlapped with dosage compensation. There are some weaknesses in each study, however. Namely, the *Sp*AID transgene killed larval males rather than the expected embryonic death[248], however the ability to recapitulate all other aspects of the phenotype suggests this is likely a byproduct of an artificial expression system rather than any evidence against it being the male-killing toxin of *Spiroplasma*. As for *wmk*, a caveat is that *Wolbachia* male killing does not naturally occur in this host, however, the comparative genomics, transgenic testing, and recapitulation of many aspects of the male-killing mechanism suggest the lack of wild type male killing in *D. melanogaster* may be due to host resistance[160]. Indeed, "hidden" male killing is common in *Wolbachia*, where a strain causes male killing in one host, but does not kill males (male killing is hidden) in another (Table 2-2)[266,341,344,368], and there is evidence

that this occurs in *D. melanogaster*[342]. It may be that the male-killing phenotype by *wmk* is therefore only possible through strong transgenic expression. In addition, the phenotype is partial, with only one third of males dying[160]. Additional work will be needed in both systems to determine if these genes are sufficient to cause male killing by the bacteria, particularly so for the more conservative candidate identified in *Wolbachia*, which does not naturally kill male hosts of a genetically tractable species as *Spiroplasma* does. This is currently technically prohibitive due to the inability to transform either microbe, but would represent a major advance if knockout of either gene in the bacteria indeed led to loss of the phenotype. Notably, much of the mechanistic work on male-killing has focused on *Spiroplasma* and *Wolbachia*, however male-killing mechanisms are not understood in other symbioses. It will be critical to focus on other cases to better understand male killing as a whole in addition to further investigating the *Wolbachia* and *Spiroplasma* systems.

The *Sp*AID and *wmk* genes are phylogenetically unrelated to the other based on BLAST searches and sequence alignments[369]. *Sp*AID has putative OTU deubiquitinase and ankyrin repeat domains, while *wmk* has two putative HTH DNA-binding domains[160,248]. Nonethless, they do share a presence on mobile genetic elements as *Sp*AID is reportedly found on a plasmid[248], while *wmk* is a prophage WO gene incorporated into the bacterial chromosome[160]. The most similar *w*Mel gene to *Sp*AID is WD0633, which is also not phylogenetically related but shares similar domains[248]. However, it did not transgenically recapitulate male killing when tested in *D. melanogaster*. Neither did any other *w*Mel homolog of *wmk* (several similar copies exist in the genome)[369]. Thus, although each gene (*Sp*AID and *wmk*) largely recapitulates its respective phenotype transgenically, they are completely distinct genes in terms of domain and phylogeny, providing further evidence for the independent origins of male killing in these two symbionts. In addition, that further testing of these additional candidates (WD0633 and *wmk* homologs) did not recapitulate male killing further supports *wmk* as the best candidate for male killing by *Wolbachia* and its prophage WO.

One curious aspect of the male killing mechanism in *Wolbachia* is the occurrence of phenotype switching and hidden male killing. In a simple scenario, the presence of a microbial gene would lead to a particular phenotype in the host. However, this is not the case with *Wolbachia* male killing. There are many examples where a strain causes male killing in one host but not in another[266,267,341,344], or where it causes male killing at one density but not another[262], or where it

causes male killing until the host develops resistance (Table 2-2)[329]. In all of these cases, a microbe capable of male killing does not exhibit the phenotype due to its environmental circumstances. Where phenotype switching occurs between hosts, it may be due to a requirement for close adaptation between toxin and host target that could be too different in another host to function properly. Where male killing is hidden due to density, it is likely because a threshold expression level is required for male killing but no other phenotypes. In cases of resistance, it is likely a random mutation in the host target that disrupts a lock-and-key model of toxin-target mechanisms. Although the reasons for phenotype loss in some cases are simple, they introduce a few cautions and questions. For example, the causative gene will be more common than the phenotype for any phenotype that can be hidden, making it difficult to identify genomes that truly lack the phenotype due to introduction of false negatives into analyses. In addition, how and why are male killing genes maintained in a system where the phenotype does not occur? Loss of the gene would be expected if it is not functioning, but hidden male killing may be maintained in another strain for many years[266]. It may therefore be that the *Wolbachia* male-killing toxin has another function or that it is not very costly to the bacteria to produce without a phenotype. Further, when male killing is lost another phenotype often arises (often CI)[368]. Is there a functional connection between them? Transgenic testing suggests no[369], but many microbes maintain both phenotypes in their arsenal[368], so there is likely some connection between them. Finally, what about other microorganisms? Phenotype switching and hidden male killing are common in *Wolbachia*, but unreported for others. The basis of this discrepancy will be an important future research question.

Applications of Male Killing

Currently, *Wolbachia* exhibiting the CI phenotype are deployed worldwide in vector control efforts. Some groups use a population suppression method where infected males are released into the wild to kill offspring of wild females, while others use a population replacement method where infected females are released to outcompete wild, uninfected females[183,370,371]. In population replacement, the combined abilities of *Wolbachia* to drive itself through a population and block viruses make it a valuable tool in fighting diseases like dengue and Zika viruses[81]. Although trials thus far have been successful in reducing the spread of some diseases[183], there are worries that eventually the viruses will become resistant to *Wolbachia*[372]. In addition, this is all done without complete knowledge of how CI works on a molecular level, so it is difficult to predict

in which situations it will work best and for how long. Thus, it is crucial not only to continue studying the genetic and molecular basis of CI to be better informed in current and future vector control efforts, but also to develop alternate methods in cases where CI may eventually fail.

In particular, male killing may have uses in conjunction or as a standalone strategy to crash populations of pests or disease vectors. Population modeling suggests that using male killing alongside other control methods such as sterile insect technique (SIT, sterile males released) or CI can help speed the time to population eradication or relax the requirements to achieve eradication[373]. However, this remains to be empirically tested. Male killing could be deployed either transgenically (uninfected insects genetically transformed to contain and express a male-killing factor), para-transgenically (*Wolbachia* transformed to express a male-killing factor), or as a natural infection in a natural host. The natural infection is the most likely to be approved since there are no genetic modifications, but given the proclivity of so many species to resist male killing (Table 2-2) and the absence of male killing in many host species, it may be difficult for a male-killing strategy to be successful. One advantage of male killing as a control strategy is that the apparent specificity between a microbial strain and closely-related hosts may help reduce the chances of it spreading beyond the intended population. Thus, male-killing may be useful where an invasive species or other isolated host must be eradicated without harm to other arthropods. Indeed, this could also conceivably be useful in conservation where here are antagonistic interactions between one species and another that should be conserved. Therefore, beyond the importance to the basic biology of many arthropod species, male killing may have applications in the field that could be better informed by continued study.

For conclusions and future directions, see Chapter 7.

Acknowledgements

The authors of this work were supported by National Institutes of Health (NIH) grants R21AI133522 to S.R.B. and F31AI143152 to J.I.P.

Chapter 3

Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility

Contributing Authors: Daniel P. LePage*, Jason A. Metcalf*, Sarah R. Bordenstein, Jungmin On, Jessamyn I. Perlmutter, J. Dylan Shropshire, Emily M. Layton, Lisa J. Funkhouser-Jones, John F. Beckmann, and Seth R. Bordenstein *Co-first authors

Abstract

The genus *Wolbachia* is an archetype of maternally inherited intracellular bacteria that infect the germline of numerous invertebrate species worldwide. They can selfishly alter arthropod sex ratios and reproductive strategies to increase the proportion of the infected matriline in the population. The most common reproductive manipulation is cytoplasmic incompatibility (CI), which results in embryonic lethality in crosses between infected males and uninfected females. Females infected with the same *Wolbachia* strain rescue this lethality. Despite more than 40 years of research[374] and relevance to symbiont-induced speciation[178,375], as well as control of arbovirus vectors[81,376,377] and agricultural pests[378], the bacterial genes underlying CI remain unknown. Here, we use comparative and transgenic approaches to demonstrate that two differentially transcribed, codiverging genes in the eukaryotic association module of prophage WO[20] from *Wolbachia* strain *w*Mel recapitulate and enhance CI. Dual expression in transgenic, uninfected males of *Drosophila melanogaster* crossed to uninfected females causes embryonic lethality. Each gene additively augments embryonic lethality in infected males crossed to uninfected females. Lethality associates with embryonic defects that parallel those of wild type CI and is notably rescued by *w*Mel-infected embryos in all cases. The discovery of cytoplasmic incompatibility factor genes *cifA* and *cifB* pioneers genetic studies of prophage WO-induced reproductive manipulations and informs *Wolbachia*'s ongoing utility to control dengue and Zika transmission to humans.

Results and Discussion

We hypothesized that the genes responsible for CI (Fig. A-1a) are present in all CI-inducing *Wolbachia* strains but absent or divergent in non-CI strains; we also predicted that these genes

would be expressed in the gonads of infected insects. To elucidate CI candidates, we determined the core genome shared by the CI-inducing *Wolbachia* strains *w*Mel (from *D. melanogaster*), *w*Ri (from *D. simulans*), *w*Pip (Pel strain from *Culex pipiens*), and *w*Rec (from *D. recens*), while excluding the pan-genome of the mutualistic strain *w*Bm (from *Brugia malayi*). This yielded 113 gene families representing 161 unique *w*Mel genes (Fig. 3-1a, Table A-1). We further narrowed this list by comparing it to (i) homologs of genes previously determined by comparative genomic hybridization to be absent or divergent in the strain *w*Au[379], a non-CI strain, (ii) homologs to genes highly expressed at the RNA level in *w*VitA-infected *Nasonia vitripennis* ovaries, and (iii) homologs detected at the protein level in *w*Pip (Buckeye)-infected *C. pipiens* ovaries. We included ovarian data with the reasoning that CI genes might be generally expressed in infected reproductive tissues, or that the CI inducer and rescue genes might be the same. Remarkably, only two genes, *w*Mel locus tags WD0631 and WD0632, were shared among all four gene subsets (Fig. 3-1b, Tables A-2 to A-4). Notably, the homolog of WD0631 in the *Wolbachia* strain *w*Pip was found at the protein level in the fertilized spermathecae of infected mosquitoes, lending support to its role in reproductive manipulation[380].

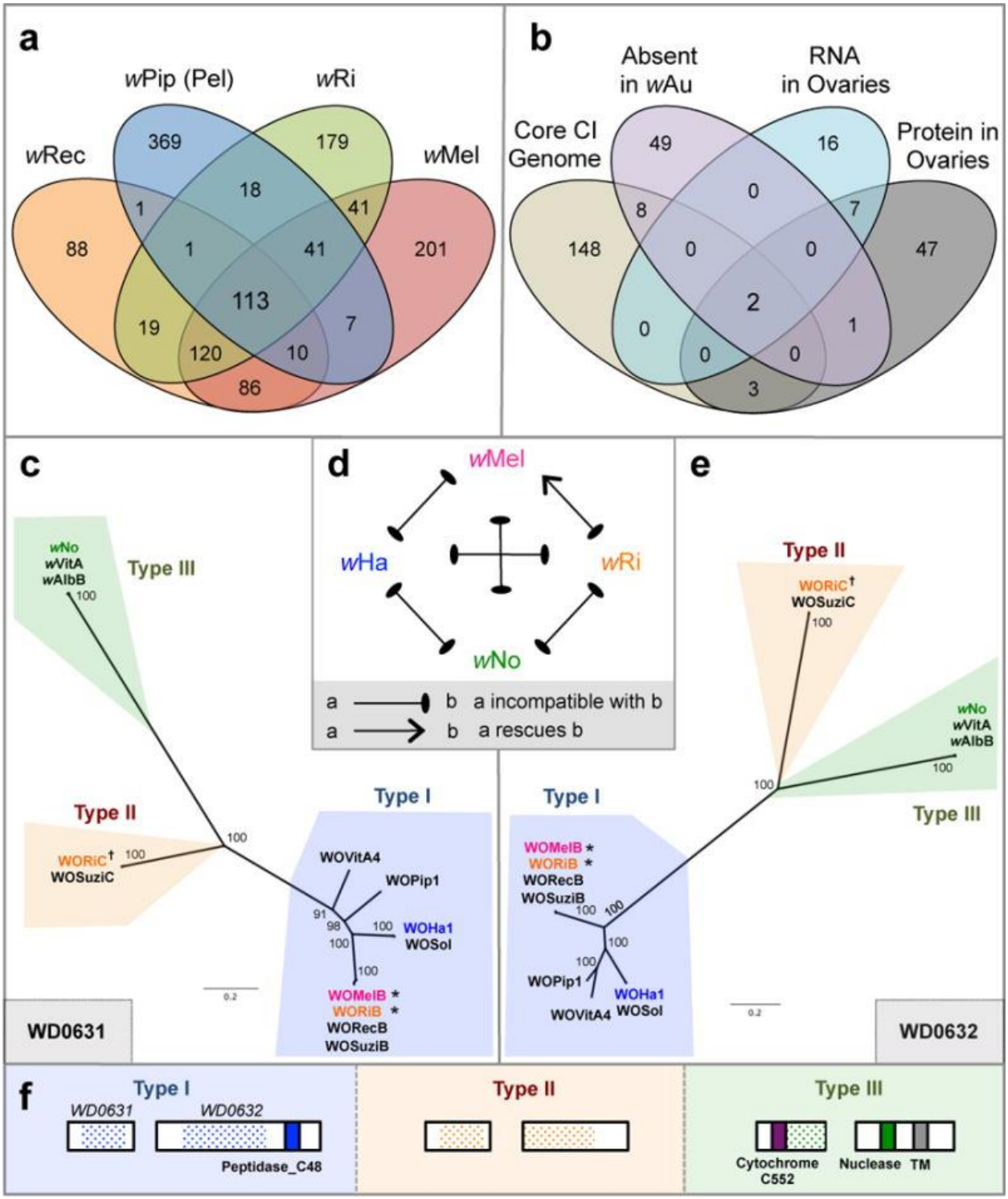


Figure 3-1. Comparative analyses reveal WD0631 and WD0632 in the eukaryotic association module of prophage WO as candidate CI genes.

(a) Venn diagram illustrating the number of unique and shared gene families from four CI-inducing *Wolbachia* strains. (b) Venn diagram illustrating the number of unique and shared *w*Mel genes matching each criteria combination. Bayesian phylogenies of (c) WD0631 (e) and WD0632 and their homologs, based on a core 256-aa alignment of WD0631 reciprocal BLASTp hits and a core

462-aa alignment of WD0632 reciprocal BLASTp hits. When multiple similar copies exist in the same strain, only one copy is shown. Consensus support values are shown at the nodes. Both trees are based on the JTT+G model of evolution and are unrooted. (d) CI patterns correlate with WD0631/WD0632 sequence homology. *w*Ri rescues *w*Mel and both share a similar set of homologs (*). The inability of *w*Mel to rescue *w*Ri correlates with a type (†) that is present in *w*Ri but absent in *w*Mel. Likewise, bidirectional incompatibility of all other crosses correlates to divergent homologs. This diagram was adapted from *Bossan et. al*[381]. (f) Protein architecture of the WD0631 and WD0632 types is conserved for each clade and is classified according to the WD0632-like domain. TM = transmembrane. Dotted shading represents the region of shared homology used to construct phylogenetic trees. For (c) and (e), the WO-prefix indicates a specific phage WO haplotype and the *w*-prefix refers to a "WO-like island," a small subset of conserved phage genes, within that specific *Wolbachia* strain.

We analyzed the evolution and predicted protein domains of these two genes and found that their homologs are always paired within the eukaryotic association module of prophage WO[20], and they codiverged into three distinct phylogenetic groups that we designate type I, II, and III (Fig. 3-1c, e, Table A-5). These relationships are not evident in the phylogeny of the *Wolbachia* cell division gene *ftsZ*, which exhibits the typical bifurcation of supergroup A and B *Wolbachia* (Fig. A-1b), or in the phylogeny of prophage WO baseplate assembly gene *gpW* (Fig. A-1c). This suggests that homologs of WD0631 and WD0632 evolve under different evolutionary pressures than genes in the core *Wolbachia* genome or in a structural module of phage WO.

Type I genes are the most prevalent amongst ten sequenced *Wolbachia* strains, and are always associated with large prophage WO regions that are often missing tail genes (Fig. A-2); it is unclear if these WO regions forge fully intact or defective interfering particles. The functions of type I WD0631 homologs are unknown, though type I WD0632 homologs contain weak homology to a putative Peptidase_C48 domain (*w*Mel, NCBI conserved domain E-value 6.69e-04, Fig. 3-1f), a key feature of Ulp1 (ubiquitin-like-specific protease) proteases[380]. Type II genes are located within more complete phage haplotypes (Fig. A-2), but the WD0632 homologs are truncated and lack recognized protein domains (Fig. 3-1f). Notably, all *Wolbachia* strains that contain type II homologs contain at least one pair of intact type I homologs. Type III genes possess WD0631 homologs with a weakly predicted cytochrome C552 domain involved in nitrate reduction (*w*No, NCBI conserved domain E-value 3.79e-03), while type III WD0632 homologs contain weak homology to the PD-(D/E)XK nuclease superfamily (*w*No, NCBI conserved domain E-value 1.15e-03) and to a transmembrane domain predicted by the transmembrane hidden Markov model[382] (Fig. 3-1f). Finally, a putative type IV group encoding a C-terminal PD-(D/E)XK

nuclease superfamily (NCBI conserved domain E-value 3.69e-03) was identified in *Wolbachia* strains *w*Pip and *w*AlbB, but not included in phylogenetic analyses because the WD0632 homologs are highly divergent (28% identity across 17% of the protein) and do not appear in reciprocal BLASTp analyses. The functions of type III and IV domains are not well understood, but a homolog of the putative nuclease domain was previously found in a selfish genetic element that mediates embryonic lethality in *Tribolium* beetles[383]. Uncertain annotations and substantial unknown sequence in these genes necessitate caution in extrapolating definitive gene functions. Importantly, the region of shared homology among the WD0632-like proteins (Fig. 3-1f) is outside the putative C-terminal Peptidase_C48 domain, suggesting that the unannotated regions could represent an ancestral CI sequence core that warrants closer inspection.

Consistent with a role in CI, the degree of relatedness and presence/absence of homologs of WD0631 and WD0632 between *Wolbachia* strains correlates with known patterns of bidirectional incompatibility (Fig. 3-1d). Among the strains *w*Ri, *w*Ha, and *w*No, only *w*Ri rescues *w*Mel-induced CI in same-species crosses[384,385]. We postulate that this is due to *w*Ri and *w*Mel sharing highly related type I homologs (99% amino acid identity), and thus likely sharing a rescue factor, while *w*Ri also has a type II homolog that may explain its ability to induce CI against *w*Mel. Meanwhile, bidirectionally incompatible pairs are highly divergent, with only 29–68% amino acid identity (Fig. 3-3a). Additionally, variation in CI strength between strains correlates with the number of copies of the WD0631/WD0632 pair (Fig. 3-3b). Strains with only one copy, such as *w*Mel, have a comparatively weak CI phenotype, while those with two or three copies, such as *w*Ri and *w*Ha, cause strong CI[385].

Given the various lines of evidence that associate these two genes with CI, we next examined the functional role of WD0631 and WD0632 in CI. For comparison, the following control genes were also used: WD0034, which is predicted to encode a PAZ (Piwi, Argonaut, and Zwille) domain containing protein (NCBI conserved domain E-value 1.85e–18), WD0508, a prophage gene annotated as a putative transcriptional regulator with two helix-turn-helix domains (NCBI conserved domain E-value 9.29e-12) in the Octomom genomic region, and WD0625, a prophage gene annotated as a DUF2466 with a JAB1/MPN/Mov34 metalloenzyme (JAMM) domain (NCBI conserved domain E-value 1.60e-41). We first examined the expression of these genes in the testes of *w*Mel-infected, one-day-old and seven-day-old *D. melanogaster* males. Since CI strength decreases significantly in aged males[386], we predicted that a CI factor would be

expressed at a lower level in seven-day-old males versus one-day-old males that both emerged on day one of the cross. Indeed, WD0631 and WD0632 show a significantly lower transcription level in aged males (Fig. 3-2). Moreover, WD0631 exhibits 18.6- and 83.0-fold higher expression than WD0632 for young and aged males, respectively (Fig. 3-2). Coupled with RNA-seq expression data[387] and operon predictor algorithms, evidence suggests that these genes are not generally acting as an operon in *w*Mel. Both prophage-encoded control genes, WD0508 and WD0625, also exhibited this age-dependent expression pattern, but the non-prophage gene WD0034 did not (Fig. 3-2). WD0640, which encodes prophage WO structural protein gpW, was also reduced in older males, suggesting that prophage genes in general are relatively downregulated in seven-day-old testes (Fig. 3-2). The phenomenon of decreased CI in older males is not due to decreases in *Wolbachia* titer over time, as the copy number of *Wolbachia groEL* relative to *D. melanogaster rp49* increases as males age, and there is no significant difference in absolute *Wolbachia* gene copies between one-day-old and seven-day-old males (Fig. A-4a,b). Since CI expression also correlates with male development time, we examined gene expression in early emerging "older brothers" (emerged on day one) and later emerging "younger brothers" (emerged on day five). Expression was statistically equivalent for WD0631 (Fig. A-4c), and slightly reduced in younger brothers for WD0632 (Fig. A-4d). These results are consistent with a small younger brother effect[388], though we did not observe a statistically significant effect on CI penetrance (Fig. A-4e).

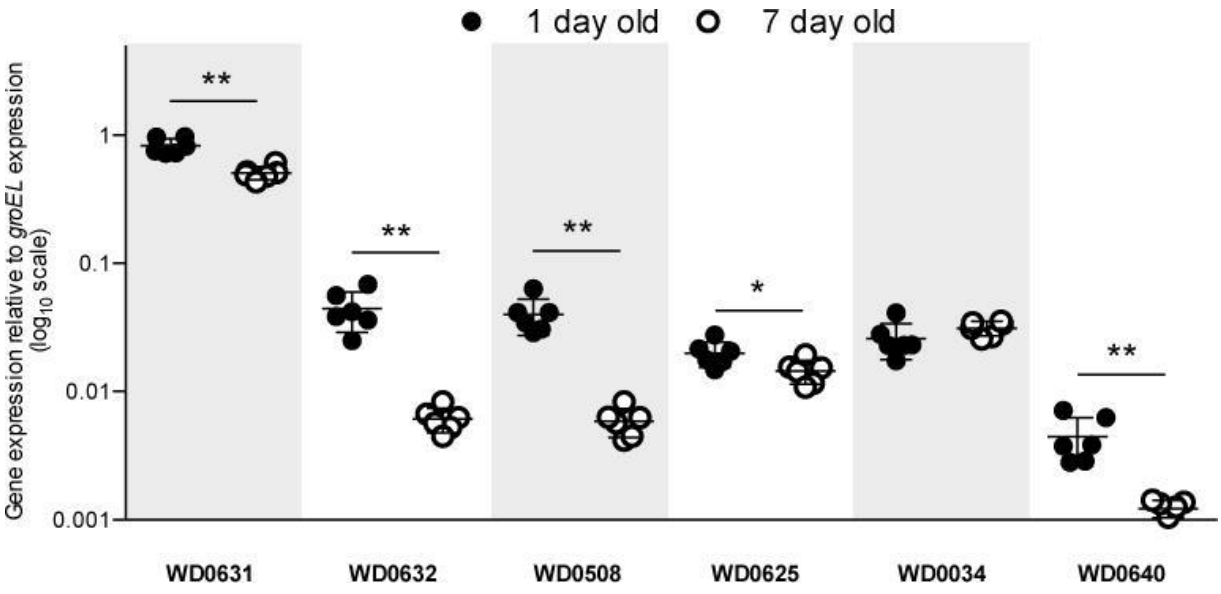


Figure 3-2. Relative expression of CI candidate and prophage WO genes decreases as males age.

RNA expression in one-day-old (1d) versus seven-day-old (7d) testes, normalized to expression of *groEL* in *w*Mel-infected *D. melanogaster* testes from the fastest developing males. Values denote $2^{-(\text{Delta Ct})}$. Bars indicate mean and standard deviation. * = P<0.05, ** = P<0.01 by Mann-Whitney U test. This experiment has been performed once. Exact p-values are provided in Table A-7.

To directly test involvement of these genes in CI, we generated transgenic *D. melanogaster* expressing candidate genes using an upstream activating sequence (UAS), since *Wolbachia* itself cannot be genetically transformed. We utilized a *nanos*-Gal4 driver line for tissue-specific expression in the germline[389,390]. We assessed CI by measuring the percentage of embryos that hatched into larvae. While wild type (WT) CI between infected males and uninfected females led to significantly reduced hatch rates, expressing each of four candidate transgenes in uninfected (fastest-developing, day one) males did not affect hatch rates when crossed to uninfected females (Fig. 3-3a, Fig. A-5a). These genes also had no effect on sex ratios (Figs. A-5b, A-6a). There are no phenotypic effects despite confirmed expression of each transgene in the testes (Figs. A-7a to A-7d).

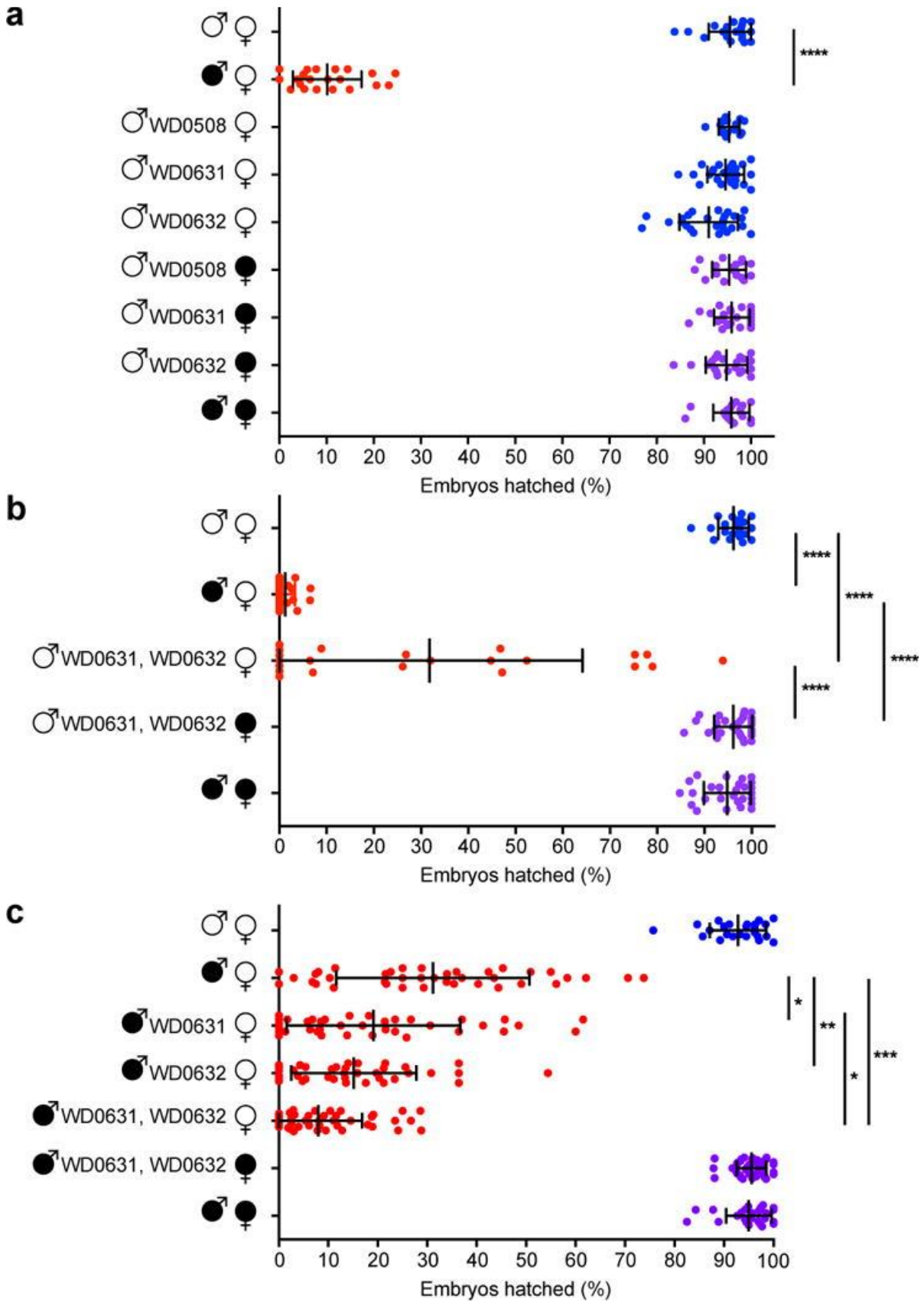

a
♂ ♀
♂ ♀
♂WD0508 ♀
♂WD0631 ♀
♂WD0632 ♀
♂WD0508 ♂
♂WD0631 ♂
♂WD0632 ♂
♂ ♂
****
0 10 20 30 40 50 60 70 80 90 100
Embryos hatched (%)
b
♂ ♀
♂ ♀
♂WD0631, WD0632 ♀
♂WD0631, WD0632 ♂
♂ ♂
****
****
****
****
0 10 20 30 40 50 60 70 80 90 100
Embryos hatched (%)
c
♂ ♀
♂ ♀
♂WD0631 ♀
♂WD0632 ♀
♂WD0631, WD0632 ♀
♂WD0631, WD0632 ♂
♂ ♂
*
**
*
***
0 10 20 30 40 50 60 70 80 90 100
Embryos hatched (%)

Figure 3-3. Dual expression of WD0631 (*cifA*) and WD0632 (*cifB*) is necessary to induce CI-like defects.
Hatch rate assays used the fastest developing males that were aged either (a, b) one-day or (c) two- to four-days in parental crosses. Parental infection status is designated with filled symbols for *w*Mel-infected parents or open symbols for uninfected parents. Transgenic flies are labeled with their transgene to the right of their male/female symbol. Unlabeled symbols represent WT flies. Data points are colored according to the type of cross, with blue indicating no CI, red for CI crosses, and purple for rescue crosses with *w*Mel-infected females. Bars indicate mean and standard deviation. * = P<0.05, ** = P<0.01, *** = P<0.001, **** = P<0.0001 by ANOVA with Kruskal-Wallis test and Dunn's multiple test correction. Statistical comparisons are between all groups (panels a and b); or between CI crosses only (panel c). All experiments were performed at least twice, except for the increase of wild-type CI by WD0631, which has been performed once. Exact p-values are provided in Table A-7.

As WD0631 and WD0632 are adjacent, coevolving genes, we reasoned that dual expression of WD0631 and WD0632 might be required to induce CI. Indeed, expression of both transgenes in the same males significantly reduced hatch rates by 68% compared to uninfected WT crosses (Fig. 3-3b), with no effect on sex ratios (Fig. A-6b). Roughly half of the crosses with transgenic males yielded hatch rates within the range observed in WT CI (3.8 ± 5.6% hatch rate). Interestingly, there was a strong positive correlation between hatch rate and clutch size when both transgenes were expressed (r_s=0.7; p=0.0003), but not in WT CI, suggesting that dilution of transgene products across larger clutches may explain variation in transgene-induced CI. It is also possible that full transgene induction of CI requires other factors, or that transgenes are not expressed at the ideal time or place for complete CI, though transgene expression in adult testes was confirmed (Fig. A-7c,d). Importantly, transgene-induced lethality is fully rescued by *w*Mel-infected females (Fig. 3-3b), indicating that these genes produce probable CI factors rather than artifacts that reduce hatch rates through off target effects that would not be rescued. We therefore name these genes cytoplasmic incompatibility factors, *cifA* and *cifB*, for WD0631 and WD0632, respectively. Type II, III, and IV homologs are designated *cif-like* until experimental evidence demonstrates they recapitulate CI.

To test if WD0631 and WD0632 transgenes act additively with *Wolbachia* to enhance WT CI levels, *w*Mel-infected male flies expressing either transgene were aged two to four days to lower WT CI penetrance before crossing with uninfected females. In support of transgene-induced enhancement of CI, hatch rates in these aged males decreased significantly compared to WT CI crosses of the same age (Fig. 3-3c), with no effect on sex ratios (Fig. A-6c). In this context, wherein aged flies cause a weaker level of WT CI, the transgenes appear to add to the quantity of CI

effectors in *w*Mel-infected tissues, causing stronger CI overall. This effect was not observed when control transgenes WD508 or WD0625 were expressed individually in *w*Mel-infected males (Fig. A-8a,b). Dual expression of WD0631 and WD0632 in *w*Mel-infected flies reduced hatch rates further than either gene alone, yet was still fully rescued by *w*Mel-infected females (Fig. 3-3c). Adding WD0625 to WD0632 in aged *w*Mel-infected males did not increase CI beyond WD0632 alone (Fig. A-8b), and had no effect on hatch rates in one-day-old uninfected males (Fig. A-8c). Finally, none of these gene combinations affected offspring sex ratios (Fig. A-9). Taken together, these findings support the conclusion that both WD0631 and WD0632 are necessary to induce the CI phenotype, and that they do not represent an artifact of the transgenic system.

To rule out the possibility that transgene-induced enhancement of CI in infected lines is due to increased *Wolbachia* titers, we quantitated amplicons of single copy genes from *Wolbachia* and *D. melanogaster*. Although there were some differences in *Wolbachia* titers between infected transgenic lines (Figs. A-10a to A-10c), the variation did not correlate with induction or magnitude of CI, signifying that decreased offspring viability was due to the direct effect of transgenes rather than changes in *Wolbachia* density. Most notably, densities significantly increased in infected flies expressing the control Octomom transgene WD0508 (Fig. A-10a) that did not enhance CI (Fig. A-8a).

Next, we tested if transgene-induced CI associates with canonical cytological defects observed in *Wolbachia*-induced CI. Although CI is typically thought to cause failure of the first mitotic division[391,392], nearly half of the embryonic arrest in incompatible crosses occurs during advanced developmental stages in *Drosophila simulans*[260,393], *Aedes polinesiensis*[394] and *Culex pipiens*[395]. We examined embryos from control and experimental crosses after one to two hours of development and binned their cytology into one of six phenotypes. While a few embryos in each cross were unfertilized (Fig. 3-4a), most embryos in compatible crosses were either in normal late-stage preblastoderm (Fig. 3-4b) or syncytial blastoderm stages (Fig. 3-4c)[396]. In WT CI, embryos exhibited three defects: arrest of cellular division after two to three mitotic divisions (Fig. 3-4d), later stage arrest associated with chromatin bridging, as is classically associated with strong CI in *D. simulans*[392] (Fig. 3-4e), or arrest associated with regional failure of division in one embryo region (Fig. 3-4f). After blindly scoring embryo cytology, we determined that arrest phenotypes (Figs. 3-d, 3-e, and 3-f) were significantly more common in the offspring of dual WD0631/WD0632 transgenic males mated to uninfected females, but these abnormalities were

rescued in embryos from *w*Mel-infected females (Fig. 3-4g). These effects were not seen with control gene WD0508 or with singular expression of WD0631 or WD0632 (Fig. 3-4h). These data again validate that transgene-induced CI, as measured through cytological defects, recapitulates WT CI. Most of the embryos arrest after two to three mitotic divisions.

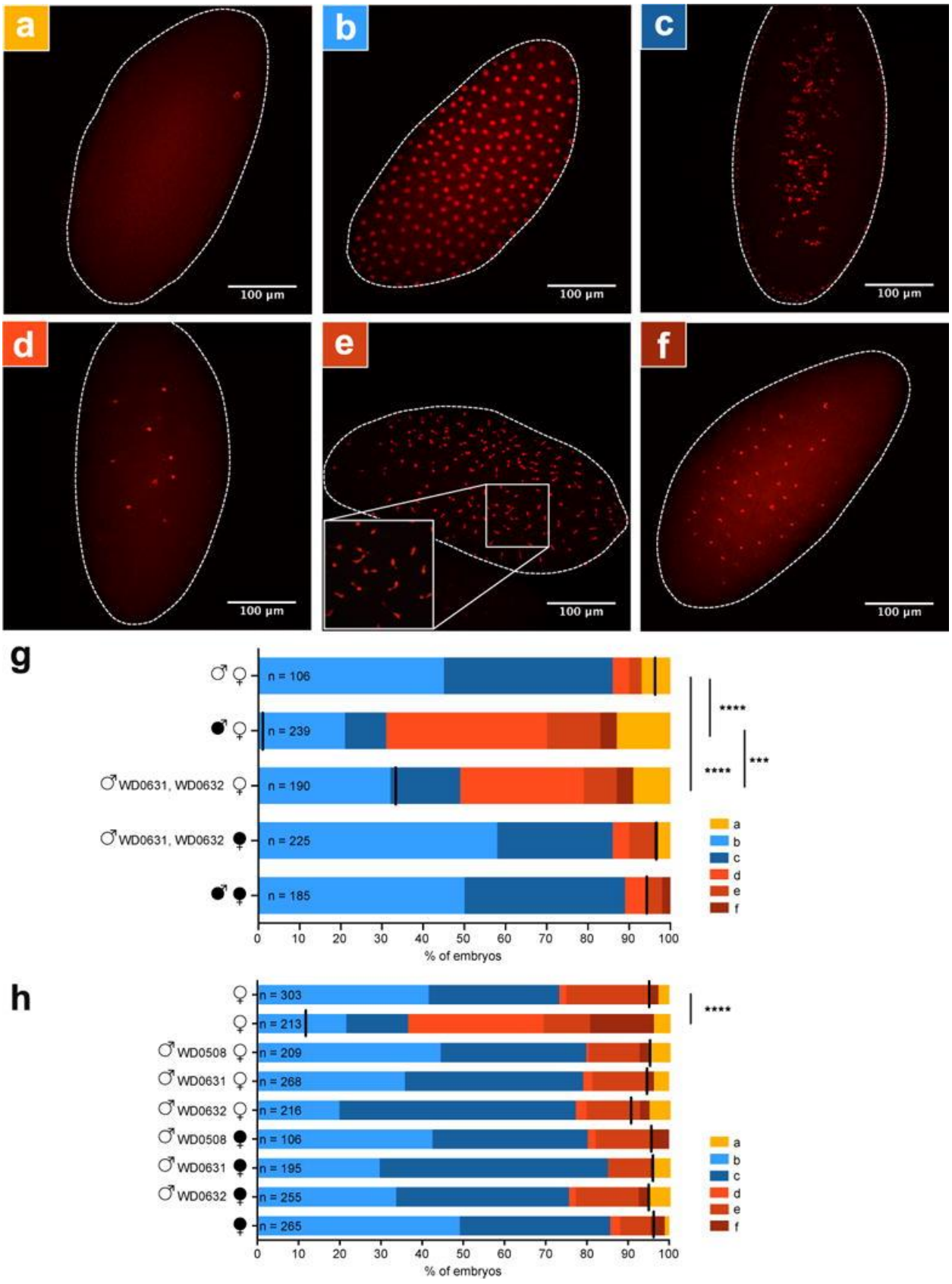

a
b
c
d
e
f
100 µm
g
♂ ♀
n = 106
♂ ♀
n = 239
♂ WD0631, WD0632 ♀
n = 190
♂ WD0631, WD0632 ♂
n = 225
♂ ♂
n = 185
****
****
***
a
b
c
d
e
f
0 10 20 30 40 50 60 70 80 90 100
% of embryos
h
♀
n = 303
♀
n = 213
♂ WD0508 ♀
n = 209
♂ WD0631 ♀
n = 268
♂ WD0632 ♀
n = 216
♂ WD0508 ♂
n = 106
♂ WD0631 ♂
n = 195
♂ WD0632 ♂
n = 255
♂
n = 265
****
a
b
c
d
e
f
0 10 20 30 40 50 60 70 80 90 100
% of embryos

Figure 3-4. Dual expression of WD0631 (*cifA*) and WD0632 (*cifB*) recapitulates CI-associated embryonic defects.
Representative embryo cytology is shown for (a) unfertilized embryos, (b) normal multi-nucleated embryos at one hour of development, (c) normal embryos near two hours of development in which nuclei begin to migrate to the periphery of the cytoplasm, and three different mitotic abnormalities: (d) failure of nuclear division after two to three mitoses, (e) chromatin bridging, and (f) regional mitotic failure. (g) The number of embryos with each cytological phenotype resulting from the indicated crosses. Infection status is designated with filled symbols for *w*Mel-infected parents or open symbols for uninfected parents. Transgenic flies are labeled with their transgene to the right of their male/female symbol. Unlabeled symbols represent WT flies. Black lines on each graph indicates mean hatch rate for the cross. *** = P<0.001, **** = P<0.0001 by two-tailed Fisher's exact test comparing normal (phenotypes b and c) to abnormal (phenotypes a, d, e, and f) for each cross. (h) Quantitation of cytological defects in control crosses. Cytology for (g) has been performed twice and cytology for (h) has been performed once. Exact p-values are provided in Table A-7.

This study identifies, for the first time, two differentially transcribed and codiverging prophage WO genes that recapitulate and enhance CI. These rapidly evolving genes are not chromosomal *Wolbachia* genes per se, but rather occur widely in the eukaryotic association module of prophage WO[20]. This module notably contains genes with amino acid sequence homologous to eukaryotes or annotated to interact with animal cells, though WD0631 and WD0632 do not appear to have eukaryotic homology. CI can therefore be categorized as a prophage WO-induced phenotype rather than a *Wolbachia*-induced phenotype. We name the genes and close homologs "cytoplasmic incompatibility factors", *cifA* and *cifB*, for WD0631 and WD0632, respectively. The *cif* name is conservatively grounded in phenotype and makes no assumptions regarding mechanism, which is notable because there are unannotated core regions throughout the *cif* genes that could have as much bearing on mechanism as the annotated domains.

The discovery of *cifA* and *cifB* genes that functionally recapitulate and enhance CI is the first inroad in solving the genetic basis of reproductive parasitism, a phenomenon induced worldwide in potentially millions of arthropod species[397]. These prophage WO genes have implications for microbe-assisted speciation, because they likely underlie CI-induced hybrid lethality observed between closely related species of *Nasonia* and *Drosophila*[175,181]. Finally, *cifA* and *cifB* are important for arthropod pest and vector control strategies, as they could be an alternative or adjunct to current *Wolbachia*-based paradigms aimed at controlling agricultural pests or curbing arthropod-borne transmission of infectious diseases[81,376-378].

Materials and Methods

Comparative genomics and transcriptomics

MicroScope[398] was used to select the set of genes comprising the core genomes of CI-inducing *Wolbachia* strains *w*Mel [NC_002978.6][399], *w*Ri [NC_012416.1][400], *w*Pip (Pel) [NC_010981.1][401], and the recently sequenced *w*Rec [GCA_000742435.1][402], while excluding the pan-genome of the mutualistic strain *w*Bm [NC_006833.1][403], using cutoffs of 50% amino acid identity and 80% alignment coverage. For the "absent in *w*Au" criterion, *w*Au microarray data were obtained from the original authors[379] and genes that were present in CI-inducing strains *w*Ri and *w*Sim but absent or divergent in the non-CI strain *w*Au were selected.

For ovarian transcriptomics, one-day old females from *w*VitA infected-*Nasonia vitripennis* 12.1 were hosted as virgins on *Sarcophaga bullata* pupae[404] for 48 hours to stimulate feeding and oogenesis. Females were then dissected in RNase-free 1× PBS buffer, and their ovaries were immediately transferred to RNase-free microcentrifuge tubes in liquid nitrogen. Fifty ovaries were pooled for each of three biological replicates. Ovaries were manually homogenized with RNase-free pestles, and their RNA was extracted using the RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol for purification of total RNA from animal tissues. After RNA purification, samples were treated with RQ1 RNase-free DNase (Promega), and ethanol precipitation was performed. PCR of RNA samples with *Nasonia* primers NvS6KQTF4 and NVS6KQTR4[405] confirmed that all samples were free of DNA contamination. RNA concentrations were measured with a Qubit 2.0 Fluorometer (Life Technologies) using the RNA HS Assay kit (Life Technologies), and approximately 5 μg of total RNA from each sample was used as input for the MICROB*Enrich* Kit (Ambion) in order to enrich for *Wolbachia* RNA in the samples. Bacterial-enriched RNA was then ethanol-precipitated, and rRNA was depleted from the samples using the Ribo-Zero Magnetic kit (Illumina) according to manufacturer's protocol. Approximately 1.5 μg of enriched, rRNA-depleted RNA for each replicate was shipped to the University of Rochester Genomics Research Center for sequencing. Library preparation was performed using the Illumina ScriptSeq v2 RNA-Seq Library Preparation kit, and all samples were run multiplexed on a single lane of the Illumina HiSeq2500 (single-end, 100 bp reads). Raw sequence reads were trimmed and mapped to the *w*VitA genome (PRJNA213627) in CLC Genomics Workbench 8.5.1 using a minimum length fraction of 0.9, a minimum similarity fraction of 0.8, and allowing one gene hit per read. With all three replicates combined, a total of 364,765 reads out of 41,894,651 (0.87%)

mapped to the *w*VitA genome with the remaining reads mapping to the *N. vitripennis* host genome (GCF_000002325.3). All *Wolbachia* genes with greater than or equal to five RNA-seq reads, with the exception of the 16S and 23S RNA genes, were selected. For non-*w*Mel data sets, the closest homologs in *w*Mel were found using BLASTp in Geneious Pro v5.5.6[406].

Protein extraction and mass spectrometry

Protein was extracted from *Culex pipiens* tissues as described previously[380]. Ovaries from 30 *w*Pip (Buckeye)-infected mosquitoes were dissected in 100% ethanol and collected in a 1.5 ml tube filled with 100% ethanol. Pooled tissues were sonicated at 40 mA for 10 seconds in a Kontes GE 70.1 ultrasonic processor, and trichloroacetic acid (TCA) was added to a final concentration of 10% (v/v). After centrifugation at 13,000 rpm in a microcentrifuge, pellets were washed with acetone:water (9:1), dried, and stored at −20°C. Samples were directly submitted to the University of Minnesota's Center for Mass Spectrometry and Proteomics for iTRAQ (isobaric tagging for relative and absolute quantification) analysis. Proteins were sorted according to their relative abundance as determined by the number of spectra from the single most abundant peptide. Because proteins can often produce varying amounts of detectable tryptic peptides depending upon protein size and lysine/arginine content, we counted only the single most abundant peptide for each protein. This quantification is justified by a previous report[407] showing that the two most abundant proteins are the *Wolbachia* surface protein (WSP; WP_007302328.1) and another putative membrane protein (WP0576; WP_012481859.1). Only proteins with at least three unique peptides (95% confidence) detected were reported, and using this criterion the false discovery rate was zero.

Evolutionary analyses

WD0631 (NCBI accession number AAS14330.1) and WD0632 (AAS14331.1) from *w*Mel were used as queries to perform a BLASTp search of NCBI's nonredundant (nr) protein sequence database with algorithm parameters based on a word-size of six and BLOSUM62 scoring matrix[408]. Homologs were selected based on the satisfaction of three criteria: (i) E-value $\leq 10^{-20}$, (ii) query coverage greater than 60%, and (iii) presence in fully sequenced *Wolbachia* and/or phage WO genomes. FtsZ and gpW proteins were identified for all representative *Wolbachia* and phage WO genomes, respectively. Protein alignments were performed using the MUSCLE plugin[409] in Geneious Pro v8.1.7[406]; the best models of evolution, according to the corrected Akaike Information Criteria (AICc)[410], were estimated using the ProtTest server[411]; and phylogenetic trees

were built using the MrBayes plugin[412] in Geneious. Putative functional domains were identified using NCBI's BLASTp, Wellcome Trust Sanger Institute's PFAM database[413], a transmembrane hidden Markov model[382] and EMBL's Simple Modular Architecture Research Tool (SMART)[414]. WD0631/WD0632 protein homology (% aa identity) was based on a 1:1 BLASTp analysis for each pair. Prophage/WO-like island association for each pair of genes was based on prophage regions identified in a previous study[20].

Gene expression assays and Wolbachia *titers*

For the male age effect, native expression of CI candidates was tested with RT-qPCR on replicate pools of 20 pairs of testes from the fastest developing, virgin males that were aged one day or seven days. RNA was extracted with the Qiagen RNeasy mini kit, DNase treated with TURBO DNase (Life Technologies) and cDNA was generated with Superscript III Reverse Transcriptase (Invitrogen). Primer sequences are listed in Supplementary Table 6. Quantitative PCR was performed on a Bio-Rad CFX-96 Real-Time System using iTaq Universal SYBR Green Supermix (Bio-Rad). 30 cycles of PCR were performed against positive controls (extracted DNA), negative controls (water), RNA, and cDNA with the following conditions: 95°C 2 min, 30× (95°C 15 sec, 56°C 30 sec, 72°C 30 sec), 72°C 5 min. Delta Ct values between the target gene and housekeeping gene *groEL* were used to determine relative gene expression. These experiments were performed once with multiple replicates for each condition.

For experiments on the younger brother effect, replicate pools of 20 pairs of testes were collected from the fastest developing virgin males that emerged on the first day (older brothers) or fifth day (younger brothers). Male siblings for the younger brother effect analysis were also collected concurrently for hatch rates described in the Hatch Rate Assays section by crossing the *w*Mel-infected males to 3- to 5-day-old *w*Mel-infected or uninfected females. RNA was extracted using the Direct-zol RNA MiniPrep Kit (Zymo), DNase treated with DNA-free (Ambion, Life Technologies), cDNA was generated with SuperScript VILO (Invitrogen), and RT-qPCR was run using iTaq Universal SYBR Green (Bio-Rad). Primers, PCR conditions, and analysis are the same as for the male age effect above. These experiments have been performed once with multiple replicates for each condition.

For gene expression in Fig. A-7, six pools of six pairs of testes were dissected from parents utilized in hatch rate assays from a repeat of Fig. 3-3a and Extended Fig. 3-5. In samples designated

"High CI" and "No CI", the males correspond to crosses that had lower or normal hatch rates, respectively. For all other samples, the flies utilized were chosen at random. RNA was extracted using the same method as the younger brother experiment above. 30 cycles of PCR were performed against positive controls (extracted DNA), negative controls (water), RNA, and cDNA with PCR conditions described above. Gel image size and brightness were adjusted in some cases for clarity. These experiments have been performed once.

For the *Wolbachia* titers, pools of testes were dissected from 15 males in ice-cold PBS. For Figs. A-10a to A-10c, brothers of those used in the corresponding hatch rates were utilized. DNA was extracted using the Gentra Puregene Tissue kit (Qiagen). qPCR was done as described above. Absolute quantification was achieved by comparing all experimental samples to a standard curve generated on the same plate. Primers are listed in the Table A-6. qPCR conditions: 50°C 10 min, 95°C 5 min, 40× (95°C 10 sec, 55°C 30 sec), 95°C 30 sec. To obtain a more accurate *Wolbachia*:host cell ratio, it was assumed that each host cell has two copies of *rp49* and each *Wolbachia* cell has one copy of *groEL*. These experiments have been performed once but with a sample size of eight for each condition.

Fly rearing

D. melanogaster were reared on a standard cornmeal and molasses-based media. Stocks were maintained at 25°C while virgin flies were 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. *Wolbachia* uninfected lines were generated through tetracycline treatment for three generations. Briefly, tetracycline was dissolved in ethanol and then diluted in water to a final concentration of 1mg/mL. 1mL of this solution was added to 50mL of media (final concentration of 20ug/mL). Freshly treated media was used for each generation. Infection status was confirmed with PCR using Wolb_F and Wolb_R3 primers[415], and flies were reared on untreated media for at least three additional generations to allow for mitochondrial recovery before being utilized[416].

Transgenic flies

Each CI candidate gene was cloned into the pTIGER plasmid for transformation and expression in *D. melanogaster*[417]. pTIGER, a pUASp-based vector designed for germline expression, exhibits targeted integration into the *D. melanogaster* genome using PhiC31

integrase[418] and tissue-specific, inducible expression through the Gal4-UAS system[419]. Cloning was performed using standard molecular biology techniques and plasmids were purified and Sanger-sequenced for confirmation before injection. At least 200 *D. melanogaster* embryos were injected per gene by Best Gene, Inc (Chino Hills, CA), and transformants were selected based on w+ eye color. All transgenic lines were made in the yw *D. melanogaster* background, and each was an isofemale line derived from the offspring of a single transformant. Homozygous lines were maintained when possible, or heterozygous flies were maintained when homozygous transgenics were inviable (WD0625/CyO). WD0508 and WD0631 insertion was carried out with the y^1 M(vas-int.Dm)ZH-2A w*; P(CaryP)attP40 line. WD0625 was inserted into BSC9723 with the genotype: y^1 M(vas-int.Dm)ZH-2A w*; PBac(y+-attP-3B)VK00002. WD0632 insertion was done using BSC8622 with the genotype: y^1 w^{67c23}; P(CaryP)attP2.

Hatch rate and sex ratio assays

Parental females were either infected or uninfected y^1w* flies (*w*Mel-infected or uninfected) and aged for 2–6 days before crossing. Uninfected y^1w^* flies were generated as described for transgenic lines. Parental males were created by crossing *nanos*-Gal4 virgin females (*w*Mel-infected or uninfected) with either y^1w^* or UAS-candidate gene-transgenic, uninfected males. Only the first males emerging between 0–30 hours from these crosses were used in CI assays to control for the younger-brother effect associated with CI[388]. To test if CI can be increased by transgenes, virgin, day one males were aged for 2–4 days before crossing to reduce the level of WT CI. Within experiments, care was taken to match the age of males between experimental and control crosses. 32–64 individual crosses were used for each crossing condition. The flies used were chosen at random from the desired group based on age, sex, and genotype. These sample sizes are based on previous studies of CI in *D. melanogaster* that detected significant differences between treatment groups[420].

To perform the hatch rate assays, a male and female pair was placed in an 8oz, round bottom, polypropylene *Drosophila* stock bottle. A grape juice-agar plate with a small amount of yeast mix smeared on top was placed in the bottle opening and affixed with tape. To create grape juice-agar plates, 12.5g of agar is mixed in 350mL of deionized water and autoclaved. In a separate flask, 10mL of ethanol is used to dissolve 0.25g tegosept (methyl 4-hyrdoxybenzoate). 150mL of

Welch's grape juice is added to the tegosept mix, combined with the agar, and poured into lids from 35×10mm culture dishes (CytoOne).

Hatch rate bottles were placed in a 25°C incubator overnight (~16 hours). After this initial incubation the grape plates were discarded and replaced with freshly yeasted plates. After an additional 24 hours, the adult flies were removed and frozen for expression analysis and the embryos on each plate were counted. The counting was not blinded. These plates were then incubated at 25°C for 36 hours before the number of unhatched embryos were counted. Larvae from each pair of flies were moved from these plates using a probe and placed in vials of standard fly media with one vial being used for each individual grape plate to be assayed for sex ratios at adulthood. A total of 10–20 vials were used for each cross type. Any crosses with fewer than 25 embryos laid were discarded from the hatching analysis while vials with fewer than 10 adults emerging were discarded from the sex ratio analysis. Statistical analysis and outlier removal, utilizing the ROUT method, were performed using Graphpad Prism v6 software.

Embryo cytology

Embryos were collected in a fashion similar to hatch rate assays except bottles contained 60–80 females and 15–20 males. All flies used were brothers and sisters of those used during corresponding hatch rates. Embryo collections and hatch rates were performed side-by-side. After initial mating overnight, fresh grape plates with yeast were provided and were removed after 60 minutes. The embryo-covered plates were then placed in the incubator at 25°C for a further 60 minutes to ensure each embryo was at least 1–2 hours old. Embryos were then moved to a small mesh basket and dechorionated in 50% bleach for 1–3 minutes. These were then washed in embryo wash solution (0.7% NaCl, 0.05% Triton X-100) and moved to a small vial containing ~2 mL heptane. An equal amount of methanol was added to the vial and then vigorously shaken for 15 seconds. After the embryos settled, the upper heptane layer and as much methanol as possible were removed, and the embryos were moved into ~500 uL fresh methanol in a 1.5 mL microcentrifuge tube. Embryos were stored overnight at 4°C. The old methanol was then removed and replaced with 250 uL of fresh methanol along with 750 uL of PBTA (1× PBS, 1% BSA, 0.05% Triton X-100, 0.02% sodium azide). After inverting the tube several times, the solution was removed and replaced with 500 uL PBTA. Embryos were then rehydrated for 15 minutes on a rotator at room temperature. After rehydrating, the PBTA was replaced with 100 uL of a 10 mg/mL RNase A

(Clontech Labs, Inc) solution and incubated at 37°C for 2 hours. The RNase was then removed and embryos were washed several times with PBS followed by a final wash with PBS-Azide (1× PBS, 0.02% sodium azide). After removing the PBS-Azide, embryos were mounted on glass slides with ProLong Diamond Antifade (Life Technologies) spiked with propidium iodide (Sigma-Aldrich) to a final concentration of 1 μg/mL. Imaging was performed at the Vanderbilt Cell Imaging Shared Resource using a Zeiss LSM 510 META inverted confocal microscope. All scores were performed blind (researcher was not aware of which slide represented which cross) and image analysis was done using ImageJ software[421]. Matched scoring, where embryos are derived from a side-by-side hatch rate, has been performed once for conditions shown in Fig. 3-4h and twice for Fig. 3-4g.

Statistical analyses

All statistical analyses were performed using GraphPad Prism software (either Prism 6 or online tools). When comparing gene expression levels or *Wolbachia* titers between two sets of data, we used a two-tailed, non-parametric Mann-Whitney U test since it does not require a normal distribution of the data. For comparisons between more than two data sets, we used the non-parametric Kruskal-Wallis one-way analysis of variance test that, if significant, was followed by a Dunn's test of multiple comparisons. This allowed for robust testing between all data groups while avoiding multiple test bias. For the cytology studies, embryos were classified as either "normal" or "CI-like" in a 2×2 contingency table, and statistical differences between the groups were calculated using a Fisher's Exact Test.

Data availability

*w*VitA transcriptome data is deposited in the Sequence Reads Archives with Bioproject PRJNA319204 and BioSample SAMN04881412. *w*Pip-infected ovarian proteome data was deposited at the Proteome Xchange Consortium via the PRIDE[422] partner repository with the dataset identifier PXD004047. All other source data is available as supplemental information with this publication.

Acknowledgments

This work was supported by NIH R21 HD086833 and NSF IOS 1456778 to Seth R.B., NSF DEB-1501398 and NIH 5T32GM008554 training grant support to D.P.L., NIH

T32GM07347 training grant support for J.A.M. to the Vanderbilt Medical Scientist Training Program, and NIH AI081322 to A.M.F. Imaging was performed in part through the use of the VUMC Cell Imaging Shared Resource (supported by NIH grants CA68485, DK20593, DK58404, DK59637, and EY08126). We thank Dr. Kristin Jernigan and Paul Snider for help with preliminary studies and Andrew Brooks for assistance with figure preparation.

Author Contributions

D.P.L. performed gene expression and hatch rate assays, embryo cytology, and assayed for transgene and infection status of flies. J.A.M. performed comparative genomics analyses, generated transgenic flies, and drafted the manuscript. Sarah R.B. performed evolutionary and bioinformatic analyses. J.O. performed hatch rates, assayed sex ratios, collected flies for all experiments, and assayed for transgene and infection status of flies. J.I.P conducted younger brother effect experiments and performed embryo cytology. J.D.S performed hatch rate assays, collected flies for parallel embryo cytology, and assayed for transgene and infection status of flies. E.M.L collected flies and performed hatch rate assays. L.J.F. obtained the *w*VitA transcriptome. J.F.B. obtained the *w*Pip proteome. Seth R.B. supervised the work and contributed to all experimental designs, data analysis, and data interpretation. All authors participated in manuscript preparation, editing, and final approval.

Chapter 4

The phage gene *wmk* is a candidate for male killing by a bacterial endosymbiont

Contributing Authors: Jessamyn I. Perlmutter, Sarah R. Bordenstein, Robert L. Unckless, Daniel P. LePage, Jason A. Metcalf, Tom Hill, Julien Martinez, Francis M. Jiggins, and Seth R. Bordenstein

Abstract

Wolbachia are the most widespread maternally-transmitted bacteria in the animal kingdom. Their global spread in arthropods and varied impacts on animal physiology, evolution, and vector control are in part due to parasitic drive systems that enhance the fitness of infected females, the transmitting sex of *Wolbachia*. Male killing is one common drive mechanism wherein the sons of infected females are selectively killed. Despite decades of research, the gene(s) underlying *Wolbachia*-induced male killing remain unknown. Here using comparative genomic, transgenic, and cytological approaches in fruit flies, we identify a candidate gene in the eukaryotic association module of *Wolbachia* prophage WO, termed *WO-mediated killing* (*wmk*), which transgenically causes male-specific lethality during early embryogenesis and cytological defects typical of the pathology of male killing. The discovery of *wmk* establishes new hypotheses for the potential role of phage genes in sex-specific lethality, including the control of arthropod pests and vectors.

Summary

Male killing is an adaptive trait for bacteria that are maternally transmitted through host populations. Such bacteria are common in arthropods and resultantly have significant impacts on host population size, mating strategy, and evolution. Moreover, male killing bacteria are under recent scrutiny as a symbiotic strategy for arthropod pest and vector control. Despite decades of research, the microbial genetic basis of *Wolbachia*-induced male killing remains elusive. Here we demonstrate that a single gene from the eukaryotic association module in prophage WO of *Wolbachia* is a candidate for male killing as it recapitulates many aspects of the phenotype when transgenically expressed in fruit flies. This discovery represents a step forward in understanding new roles of phage WO genes in shaping arthropod hosts and may inform the potential use of male killing in worldwide pest and vector control strategies.

Introduction

Wolbachia (order Rickettsiales) infect an estimated 40-52% of all arthropod species[189,397] and 47% of filarial nematode species[423], making them the most widespread intracellular bacterial symbiont in animals. Concentrated in host testes and ovaries, *Wolbachia* primarily transmit cytoplasmically from mother to offspring[68,114]. In arthropod reproductive tissues and embryos, *Wolbachia* deploy cunning manipulations to achieve a greater proportion of transmitting females in the host population. Collectively, these strategies are categorized as reproductive parasitism.

Male killing, or selective death of an infected female's sons[245], is one such form of reproductive parasitism[262,336]. It enhances the fitness of *Wolbachia*-infected females in three potential ways: (i) reducing brother-sister competition for limited resources[335], (ii) reducing inbreeding[277], and/or (iii) providing nutrients in cases where infected sisters cannibalize embryos of their dead brothers[277]. Male-killing *Wolbachia* are widespread in several major insect orders[176] and in pseudoscorpions[26]. In addition, male-killing *Spiroplasma*[358], *Rickettsia*[277], and *Arsenophonus*[424] occur in diverse hosts including flies[358], ladybugs[277], and wasps[424].

Male killing can have several significant impacts on host evolution[175,178,181,345]. For example, male death may lead to host extinction or reduce the effective population size of the host. As a consequence, theory specifies that fixation of deleterious alleles in host populations is more likely, and fixation of beneficial alleles is conversely less likely[327,425]. Male killing can also impose strong selection on hosts to counter the sex ratios shifts and lethality[345]. Evolutionary outcomes include mate preference between uninfected males and females[176], a shift towards more mate-attracting behaviors by females or male mate choice[176], and suppression of the phenotype[266,280,329,345].

As they manipulate arthropod reproduction to drive through host populations, *Wolbachia* are currently deployed in two vector control strategies: population suppression to reduce the population size of mosquitoes, and population replacement to transform mosquito populations that transmit pathogens to ones that cannot transmit pathogens[81,426]. In these cases, mosquitoes are released with *Wolbachia* that cause cytoplasmic incompatibility (CI), in which offspring die in crosses between infected males and uninfected females. Notably, population genetic modeling demonstrates that male killing can be deployed in conjunction with population suppression techniques to speed up eradication or reduction of a target arthropod population and increase the likelihood of success[373]. However, the genetic basis of *Wolbachia* male-killing has remained a

mystery for more than sixty years[427] and the causative gene of the *Spiroplasma* male-killing phenotype has only recently been reported[428]. Thus, potential vector and pest control applications of male killing have yet to be experimentally validated.

In this study, we sought to determine the genetic basis of the male-killing phenotype in *Wolbachia*. Our previous comparative genomic, transcriptomic, and proteomic analyses identified two prophage WO genes, *cifA* and *cifB,* that underpin the induction and rescue of CI by *w*Mel *Wolbachia* in *D. melanogaster*[158,161]. *cifA* and *cifB* reside in the newly characterized eukaryotic association module of prophage WO that is enriched with many sequences predicted to have eukaryotic functions and homologies[20,158,429]. Building on this previous analysis, we pursued characterization of genes that may also be responsible for male killing. Notably, *Wolbachia* can be multipotent because some strains induce multiple reproductive parasitism phenotypes (e.g., CI and male killing) depending on the host background or environmental conditions[262,265,266,341]. For example, the *w*Rec strain of *D. recens* causes CI in its native host, but it kills males when introgressed into the genetic background of its sister species, *D. subquinaria*[266]. Importantly, *w*Mel and *w*Rec share 99.7% nucleotide identity[402], which raises the hypothesis that the CI-inducing *w*Mel genome may also harbor male-killing genes.

A long-standing question is whether multipotency is due to pleiotropy of the same gene(s) expressing different reproductive parasitism phenotypes or alternatively if different genes underpin the various forms of reproductive parasitism. We previously assessed several reproductive parasitism gene candidates in *w*Mel *Wolbachia* for both male killing and CI, including *cifA* and *cifB*, and we ruled out their involvement in male killing[158]. However, other genes may still be involved. Although *w*Mel is not known to naturally cause male killing, it is of interest because it is the native strain of the only host that is genetically tractable and is closely-related to a natural male killer, making it a useful system to test gene candidates for the phenotype.

There are several expectations for a putative *Wolbachia* male-killing gene. First, we expect transgenic expression will recapitulate the embryonic cytological defects typically induced by male killing[356]. Second, native expression of the candidate gene will occur by the time male death naturally occurs in a given host[266,356]. Third, a male-killing gene would be shared across male-killing strains in *Wolbachia* but not necessarily absent from strains unknown to cause male killing. In other words, the gene may be more common than the phenotype because hosts frequently develop resistance to male killing, presumably due to the strong evolutionary pressure to avoid

extinction[266,328,329,345,430]. As previously mentioned, *Wolbachia* can induce either male killing or CI in different hosts or rearing conditions[262,266,329,341], which may be related to resistance in some hosts. Fourth, if there is a single gene that causes male killing in most or all cases, then the gene may rapidly evolve due to natural selection in diverse host backgrounds that suppress male killing. Here, based on genomic analyses, transgenic expression, and cytological characterizations in *Drosophila melanogaster* infected or uninfected by *w*Mel *Wolbachia*, we report the discovery of a gene in the eukaryotic association module of prophage WO that is a candidate for male killing.

Results

Genomic analysis of male-killing gene candidates

To generate a shortlist of male-killing gene candidates, we used the following criteria and assumptions: (i) universal presence in the genomes of male-killing strains *w*Bif from *D. bifasciata*[427], *w*Inn from *D. innubila*[336], *w*Bor from *D. borealis*[431], and *w*Rec from *D. recens*[266]; (ii) genomic location in prophage WO because parasitic *Wolbachia* all have intact or remnant prophage WO regions with eukaryotic association module genes[20]; notably, the two previous parasitism genes, *cifA* and *cifB*, are both in this module of prophage WO, making it likely that other parasitism genes share a similar origin; (iii) exclusion of highly repetitive elements, including insertion sequence elements, reverse transcriptases of group II intron origin, and large serine recombinases that likely facilitate phage WO lysogeny; and (iv) exclusion of disrupted genes (e.g., early stop codons) in one or more strains (Table B-1 for list of excluded genes).

Table 4-1 shows seven candidate genes that fit these criteria. One of these genes, *cifA*, was previously evaluated by transgenic expression[158], and it did not exhibit a biased sex ratio. Others include a predicted ankyrin repeat (WD0550), two Rpn genes (recombination-promoting nucleases WD0297, WD0627), Phospholipase D (WD1243), and a hypothetical protein (WD0628). The remaining gene, WD0626, was identified in the previous multi-omic analysis that uncovered the *cif* genes[158]. This candidate gene, hereafter denoted *wmk* for *WO-mediated killing*, is a putative transcriptional regulator in prophage WOMelB that is predicted to encode two helix-turn-helix (HTH), XRE family DNA-binding domains (NCBI conserved domains $E= 5.9 \times 10^{-11}$, $E= 6.5 \times 10^{-10}$). *wmk* in *w*Mel has a single amino acid difference relative to its homolog in *w*Rec. Due to the association of *wmk* with two different candidate gene analyses for reproductive parasitism and

preliminary observations that transgenic expression associated with a sex ratio bias, we further assessed it as a putative male killing gene.

Table 4-1. Comparative genomic analysis of male-killing gene candidates.
After applying all criteria in the genomic analysis, seven candidates for male killing were identified. All seven gene candidates are listed with their functional annotation and locus tags from both *w*Mel and the closely related *w*Rec strain. BLASTP results of the homologs are also shown with the percent coverage, E-value, pairwise identity, and number of nucleotides for each strain. For inclusion and exclusion criteria, see Table B-1. WD0626 from *w*Mel is the gene hereafter denoted *WO-mediated killing* or *wmk.*

Annotation	*w*Mel Locus Tag	*w*Rec Locus Tag	Ref-Seq Coverage	E-Value	Pairwise % Identity	*w*Rec	*w*Mel
Ankyrin Repeat	WD0550	wRec0541	100%	0	95%	789	990
Transcriptional Regulator	WD0626	wRec0560	100%	0	99%	912	912
Rpn (Recombination-Promoting Nuclease)	WD0627	wRec0561	100%	0	99%	897	897
Hypothetical Protein	WD0628	wRec0562	100%	0	100%	540	540
CifA (CI Component)	WD0631	wRec0566	100%	0	99%	1425	1425
Rpn (Recombination-Promoting Nuclease)	WD0296	wRec0561	81%	0	88%	897	912
Phospholipase D	WD1243	wRec1232	100%	0	99%	531	531

https://doi.org/10.1371/journal.ppat.1007936.t001

The wmk *gene is common and found in all sequenced male-killing genomes*

Phylogenetic analyses indicate that *wmk* homologs are common in phage WO-containing *Wolbachia* including the above-mentioned male-killing strains (Fig. B-1), *w*Bol from *Hypolimnas bolina* butterflies (causes CI when male killing is suppressed)[329,345], and *w*CauB from *Cadra cautella* moths (causes male killing in non-native host)[341], along with many strains not known to cause male killing (Fig. B-1a). *wmk* is in the eukaryotic association module of prophage WOMelB, resides just a few genes away from the *cif* genes, and exists in multiple divergent copies in some strains (Fig. B-1b and Fig. 4-1)[20]. Phylogenetic analyses indicate that *wmk* sequence relationships do not cluster into typical *Wolbachia* supergroups (Fig. B-1a), specifying independent evolution relative to the core *Wolbachia* genome. This finding is similar to that of other prophage WO genes including *cifA*, *cifB*, and the baseplate assembly gene, *gpW*[158]. It is attributable to the high rates of horizontal phage WO transfer between *Wolbachia* coinfections[432]. Similar to *cifA* and *cifB*[433], *wmk* homologs are notably disrupted in the parthenogenesis-inducing *Wolbachia* strains *w*Uni from *Muscidifurax uniraptor* wasps, *w*Tpre from *Trichogramma pretiosum* wasps, and *w*Fol from *Folsomia candida* springtails. The gene is also absent in the male-killing MSRO strain of *Spiroplasma poulsonii*, which contains the recently reported male-killing gene, Spaid[428]. Spaid has

OTU deubiquitinase and ankyrin repeat domains and lacks direct homologs in *Wolbachia*[428], indicating separate evolutionary origins of Spaid and *wmk*. In addition, genomic analyses suggest the full version of *wmk* in phage WO potentially originated from a fusion or duplication event with gene(s) in the non-prophage region of the *Wolbachia* chromosome. Indeed, homologs of the N-terminal XRE-family HTH domain occur in distantly related nematode *Wolbachia* strains (*w*Wb, *w*Bm, *w*Ppe) and the sister genera *Ehrlichia* (Table B-2) that all lack prophage WO.

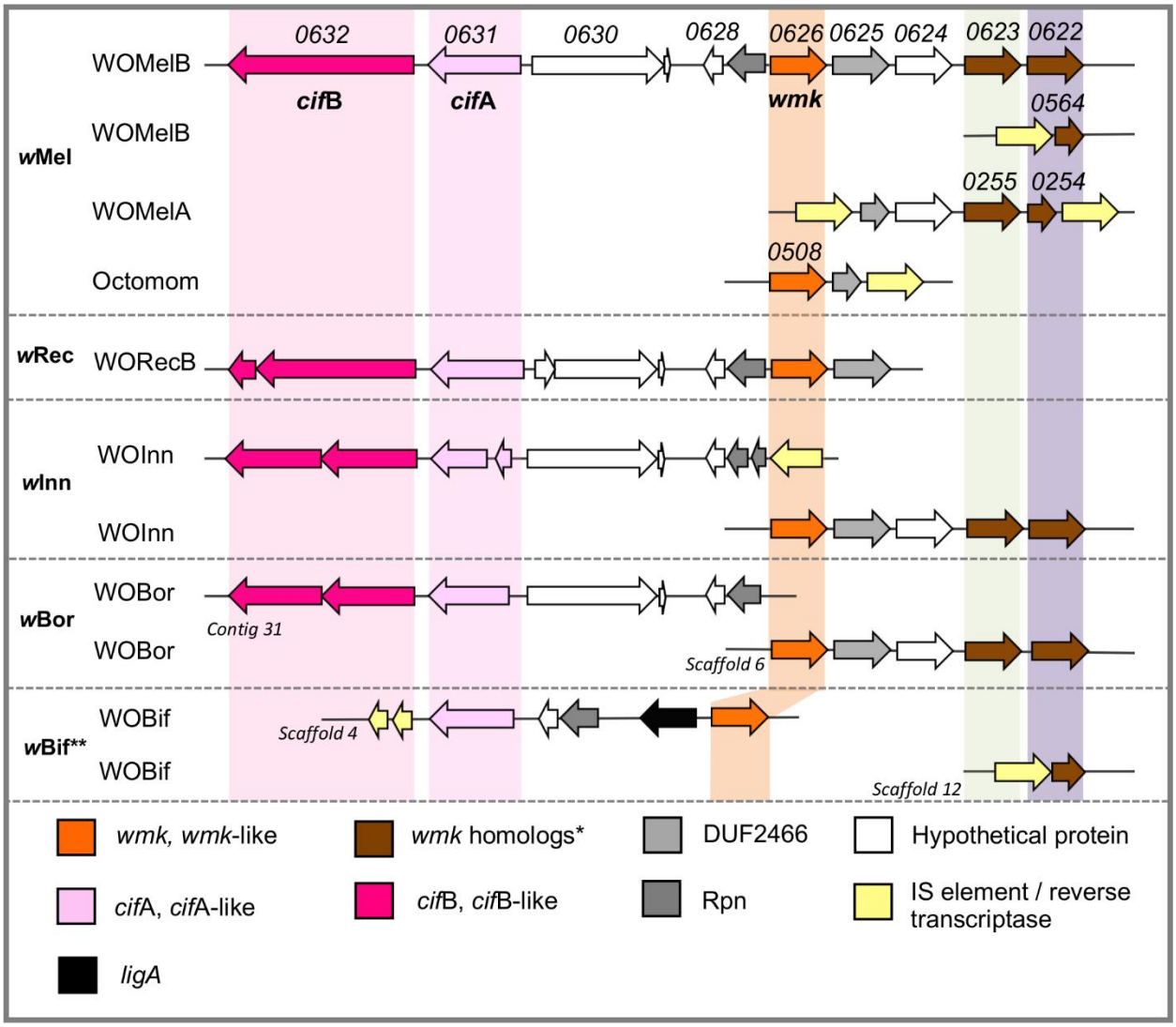


Figure 4-1. Comparative genomics of *wmk* and its homologs in *w*Mel and male-killing strains.
Prophage WO gene regions containing *wmk*, *wmk*-like homologs, and CI genes *cif*A and *cif*B are listed by *Wolbachia* strain in bold and then prophage. At least one *wmk* homolog is associated with each *Wolbachia*-induced male killing strain. Genes pointing in the same direction are on the same DNA strand. The distance between *wmk* and *cif*A is approximately 5 kb. Shading highlights

homologs in each strain. (*) *wmk* homologs are annotated as transcriptional regulators in the *Wolbachia* reference genomes and encode helix-turn-helix XRE domains (Table B-4). (**) While *w*Bif reportedly induces weak CI after temperature treatment[262], the assembled genome does not contain *cifB*.

Transgenic expression of wmk *causes a female-biased sex ratio*

To evaluate the function of *wmk*, we generated transgenic *D. melanogaster* flies that express codon-optimized *wmk* with the Gal4-UAS expression system because genetic editing of *Wolbachia* is not currently possible. We evaluated three other transgenes in a similar manner: WD0625 in prophage WO that encodes a putative MPN/Mov34/PAD-1 metalloprotease domain (DUF2466, NCBI conserved domain E= 3.85 x 10^{-41}) because it is adjacent to *wmk* and may in theory be cotranscribed with *wmk*, WD0508 in the prophage WO-associated Octomom region that is another predicted transcription regulator with two XRE-family HTH DNA-binding domains (NCBI conserved domains E= 1.70 x 10^{-9}, E= 1.99 x 10^{-11}, a homolog of *wmk*), and WD0034, a non-phage, hypothetical protein-coding gene that is hereafter labeled 'control gene' and shares a transgenic insertion site with *wmk*. These three genes do not recapitulate CI[158]. In the experiments below, all transgenes were expressed in heterozygous flies under the control of an *Act5c*-Gal4 driver, which leads to ubiquitous transgene expression beginning with zygotic transcription ~2h after egg deposition (AED). Genetic crossing schemes are described in the methods.

To assess if *wmk* causes sex-specific lethality, we first quantified adult sex ratios in gene-expressing (*Act5c*-Gal4; UAS-*wmk*) flies using a ubiquitously-expressing actin (*Act5c*) driver. *wmk* transgene expression results in a significant reduction in the average male:female sex ratio (number of males / number of females) to 0.65, or a 35% reduction in gene-expressing males (Fig. 4-2). The sex ratio is approximately 1 in wild type flies and in transgenic flies that either do not express *wmk* (CyO; UAS-*wmk*) or express a control gene (Fig. 4-2). All sex ratios represent a normal range of variance observed in previous experiments[158,428,434-436]. For example, natural male-killing *Wolbachia* strains cause variable offspring sex ratios that range from 0.5 to 0 (all females) in *D. innubila*[264,437], and 0.2 to 0 in *D. subquinaria*[266], although most cases are all female. For the three other prophage WO genes, transgenic expression in uninfected flies does not significantly change sex ratios (Figs. B-2a to B-2c), indicating the *wmk*-induced phenotype is not due to a generalized, transgenic artifact. Further, we explored whether another gene could be additionally involved. We tested dual expression of *wmk* and WD0625, as they are adjacent and could

potentially function together. Dual expression does not change the degree of male death (Figs. B-2a to B-2c), demonstrating it is not involved in the phenotype. In addition, ovarian transgene expression of *wmk* by the maternal triple driver (MTD) that loads product into developing oocytes[438] did not result in a biased sex ratio (Fig. B-2d) despite confirmed expression (Fig. B-2e). The lack of phenotype under the MTD driver is likely due to insufficient transcript levels in the embryo as MTD is a germline-specific driver expressed in mothers before eggs are laid, whereas *Act5c* is ubiquitously expressed by the embryo itself. However, transgenic expression of *wmk* via the *armadillo* driver, which expresses genes ubiquitously beginning in embryogenesis, yields sex ratios that are similar to that of the *Act5c* driver (Fig. B-3a), despite an order of magnitude reduction in expression level (Fig. B-3b). These findings indicate that expression at *Act5c* levels is not necessary to induce the phenotype, and zygotic transcription (~2 h AED) of *wmk* is required for the sex ratio effect. Thus, investigations so far have not revealed conditions that might alter the proportion of male death. Notably, this finding parallels the timing of embryonic mortality during early zygotic transcription in the *D. melanogaster* male-killer, *Spiroplasma poulsonii*, although it differs in that maternal expression does not recapitulate the *wmk* phenotype, while some aspects of the *Spiroplasma* phenotype can be recapitulated with maternal expression[428].

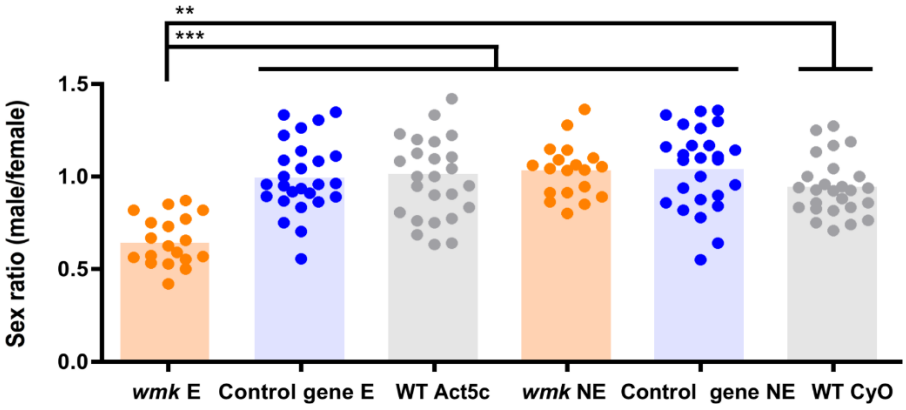


Figure 4-2. Transgenic expression of *wmk* causes a female-biased sex ratio.
Each sample point represents the adult offspring produced by a replicate family of ten mothers and two fathers (average offspring number per data point is 90). Bars represent the average sex ratio. Control gene flies have the *Wolbachia* transgene WD0034. WT is the BSC8622 strain. E = expressing, NE = non-expressing, *Act5c* has an *Act5c*-Gal4 gene, CyO has the CyO chromosome. *wmk*-expressing flies have a significantly female-biased sex ratio against all other genotypes. This experiment has been done four times. Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction. **p<0.01, *** p<0.001. Orange dots represent *wmk*, blue dots represent the control gene, and gray dots represent the WT strain.

The *wmk*-induced change in sex ratio is also not consistent with other types of reproductive parasitism for several reasons. First, CI is not known to have a sex ratio bias except in haplodiploid species[158]. Second, the male lethality phenotype and transgene expression begin long after hallmark CI defects such as delayed histone deposition in fertilized embryos[439]. Third, an infected maternal background does not rescue the *wmk* phenotype, as would be expected if the phenotype were linked to CI (Fig. B-4a). Fourth, neither *wmk* expression nor dual expression of *wmk* and WD0625, a putative partner gene due to its adjacent location, causes or rescues CI when expressed with the *nanos*-Gal4 driver used in CI experiments for germline-specific expression[158] (Figs. B-4b and B-4c). Fifth, the bias in sex ratio cannot result from genetic males developing as females (feminization) because *wmk* expression does not increase the absolute number of females compared to controls (Fig. B-4d). Finally, parthenogenesis (virgin females produce all female offspring) cannot explain the male lethality phenotype because transgenic expression occurs with a paternal chromosome present.

Transgenic expression of wmk *recapitulates embryonic death and cytological defects*

Wolbachia–induced male killing occurs either during embryogenesis or larval development in *Drosophila*[264,266,356]. Embryonic cytological defects associated with *Wolbachia* male killing begin largely at the time of host embryonic cellularization (~2.5 h after egg deposition, AED) and span abnormal nuclei distribution, chromatin bridging, and pyknosis in male embryos of *D. bifasciata*[356]. To determine if *wmk* transgene expression in *D. melanogaster* recapitulates the nature and timing of the defects, we stained DNA with propidium iodide in wild type (WT) embryos and in embryos expressing either *wmk* or the control transgene. We then monitored the defects in embryos (only half of the embryos are expected to express the transgene, see methods). Several different defects were observed (Figs. 4-3a to 4-3d). In embryos fixed 1-2 h AED, there was no significant difference in cytological defects of *wmk*-associated offspring compared to controls (Fig. 4-3i). However, in embryos fixed 3-4 h AED, cytological defects were enriched in *wmk*-associated embryos (28%) relative to control gene (11.8%) and wild type embryos (10.3%) (Fig. 4-3j). Since significantly more defects occur in embryos fixed 3-4 h AED but not in those fixed 1-2 h AED, the male lethal defects could commence between 2-4 h AED. These results also indicate that cytological defects specifically occur soon after zygotic transcription of *wmk*, as only a zygotic driver, not a maternal egg loader, is able to induce the phenotype.

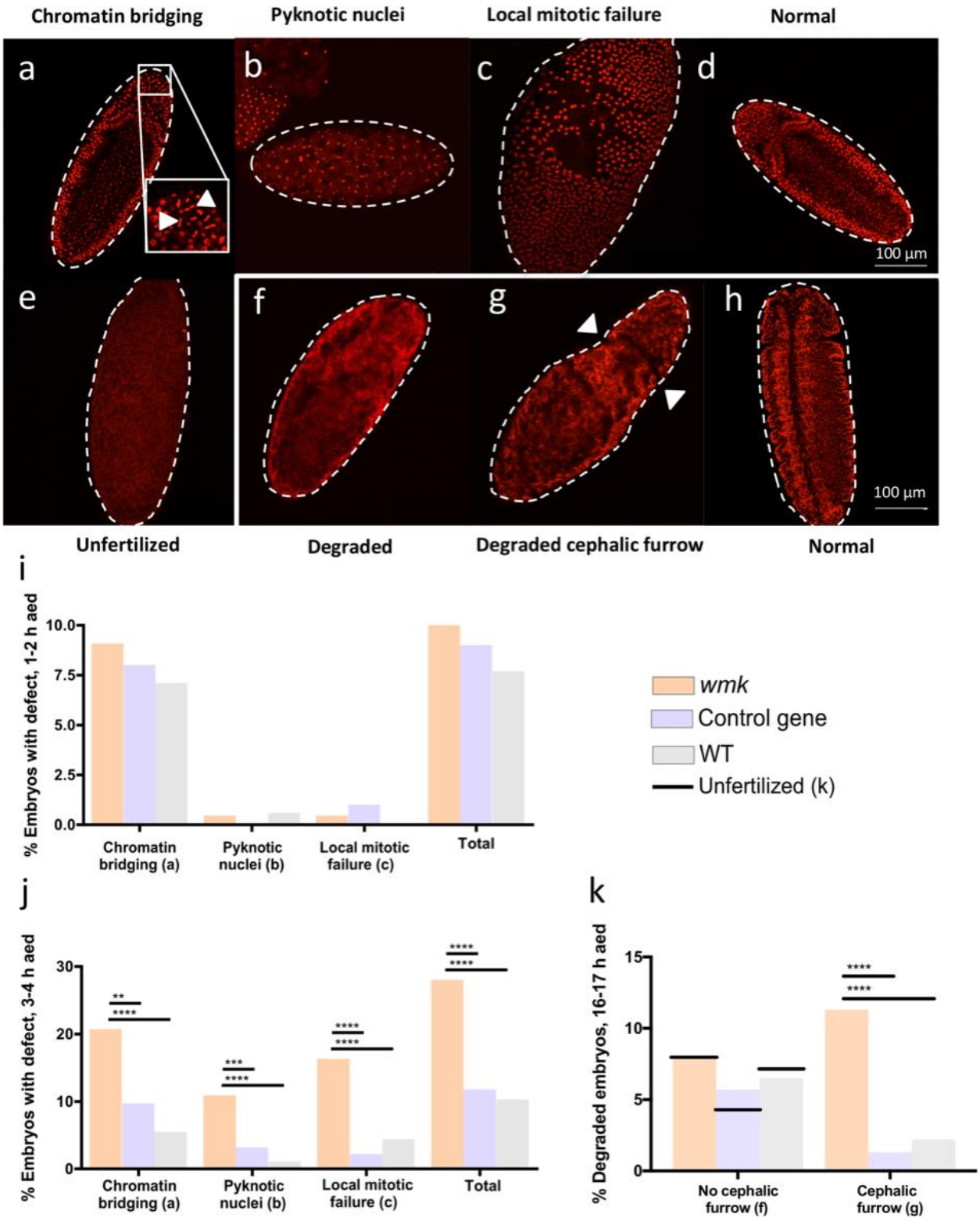


Figure 4-3. Transgenic expression of *wmk* causes cytological defects in early embryogenesis.

Data are from pooled embryos (both sexes, expressing and non-expressing) with either *wmk*, the control gene, or an uninfected wild type (WT) background (see methods). (a-c) Defective *wmk* embryos fixed 3–4 h after egg deposition (AED) exhibit either chromatin bridging (arrowheads), pyknotic nuclei, or local mitotic failure leading to gaps in the distribution of nuclei, respectively. (b) Image has been brightened for visibility. (d) Image of a normal control gene embryo fixed 3–4 h AED. (e) Image of unfertilized embryo fixed approximately 3–4 h AED. (f) Image of degraded *wmk* embryo fixed 16–17 h AED with no distinct nuclei and no visible segmentation. (g) Image of a degraded *wmk* embryo fixed 16–17 h AED with no distinct nuclei, but the cephalic furrow is (indicated by arrowheads). (f) and (g) are brightened in order to see their differences. (h) Image of normal control gene embryo fixed 16–17 h AED. (I) Graph quantitating the percentage of embryos exhibiting DNA defects that were fixed 1–2 h AED. N = 220 for the *wmk* cross, N = 200 for the control gene cross, and N = 169 for the WT cross. Total refers to the total percentage of embryos with one or more of the three defects (embryos can have more than one, as in (a)). All differences within each defect category were not statistically significant. (j) Graph of the percentage of embryos exhibiting DNA defects that were fixed 3–4 h AED for *wmk*, control gene, and WT crosses. N = 276 for the *wmk* cross, N = 273 for the WT cross, and N = 279 for the control transgene cross. (k) Graph of the percentage of degraded embryos fixed 16–17 h AED in the *wmk*, control gene, and WT crosses. N = 327 for the *wmk* cross, N = 315 for the control transgene cross, and N = 231 for the WT cross. The percent of unfertilized eggs is the expected percent given the observed rate of unfertilized sibling eggs fixed 3–4 h AED (*wmk*, 8%, N = 324; control gene, 4.5%, N = 202; WT, 7%, N = 217). Statistics for (i), (j), and (k) were performed with a Chi-square test comparing the three genotypes within each defect category. These experiments have been performed once. The white border around (f, g, & h) indicates embryos fixed 16–17 h AED, while the rest (a-e) are embryos fixed 3–4 h AED. All images were taken at 20X zoom, except the inset image in (a) that is a zoomed in image of the same region. ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$.

In *w*Bif-infected *D. bifasciata*, male embryos 15-20 h AED have several large defects including incompletely formed regions and lack of differentiation or segmentation[356]. To determine if the defects in early *wmk*-expressing embryos result in similar abnormalities later in development, we fixed sibling embryos 16-17 h AED. We discovered and assessed degraded embryos (embryos with cloudy staining from degraded DNA and lack of distinct nuclei) in *wmk*-associated offspring compared to controls. One category of degraded embryos had no visible cephalic furrow or segmentation similar to unfertilized eggs (Fig. 4-3e and 4-3f). These embryos occurred equally across all treatment groups at a low percentage similar to that of unfertilized eggs (Fig. 4-3e and 4-3k). This category likely represents decomposing, unfertilized eggs. A second degraded form exhibited a cephalic furrow that demarcates the head from the thorax (Fig. 4-3g), but it lacked other normally visible segmentation (Fig. 4-3h), similar to the lack of segmentation in infected embryos. There were approximately 10-fold more degraded embryos with a cephalic furrow in the *wmk* cross versus controls (Fig. 4-3k). This finding suggests the timing of death is

soon after the commencement of the cephalic furrow formation, which occurs at approximately 3 h AED. As noted above, it is also approximately the time point when cytological defects are first observed (Fig. 4-3j). The furrow formation is largely complete by 4 h AED, and it is visible in the degraded embryos, suggesting most embryos reach this developmental time point before death. Though this furrow phenotype is not described in natural contexts, the literature demonstrates that there are highly defective areas in embryos later in development[356]. The furrow phenotype likely occurs in transgenic individuals because of consistent, strong expression of a transgene rather than natural expression levels that may vary in individuals due to differences in *Wolbachia* titer or gene expression. However, the lack of segmentation is known in natural contexts. Interestingly, the marked number of degraded cephalic furrow *wmk* embryos is proportional to the number of missing males in adult sex ratios (Fig. 4-2). These results imply that the degraded embryos 16-17 h AED and the reduced sex ratios of surviving adults are the result of *wmk*-induced defects in early male embryos. Taken together, there are four key results: (i) *wmk* induces DNA defects 2-4 h AED, (ii) embryos arrest after cephalic furrow formation, (iii) embryos become degraded by late stages of embryogenesis, and (iv) embryonic defects lead to downstream reductions in sex ratios of surviving flies. Notably, the 2-4 h time window is when defects begin to significantly occur in *D. bifasciata*. The corresponding adult sex ratios for this experiment are shown in Fig. B-5a.

Next, we confirmed that the cytological defects in embryos 3-4 h AED are male-biased using fluorescent *in situ* hybridization (FISH) with a DNA probe specific to the Y chromosome (Fig. B-6, expressing and non-expressing embryos, see methods). 40% of male *wmk* embryos exhibit defects versus 9% of female *wmk* embryos and 9-10% of WT and control gene embryos (Fig. 4-4a). In addition, while the embryonic sex ratios are not biased at 1-2 h AED, they are biased among viable (non-degraded) embryos fixed 16-17 h AED (Fig. 4-4b), as expected. The corresponding adult sex ratio of 0.68 was similar to the embryonic sex ratio (Fig. B-5b), further indicating that male killing occurs during embryogenesis. These results specify that defects and degradation are enriched in males.

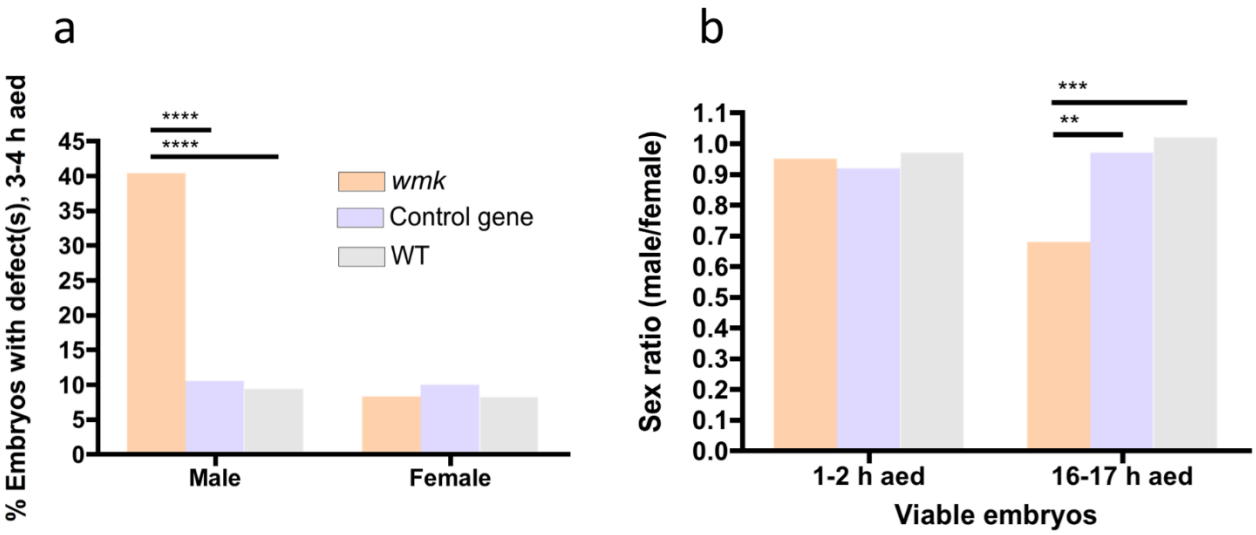
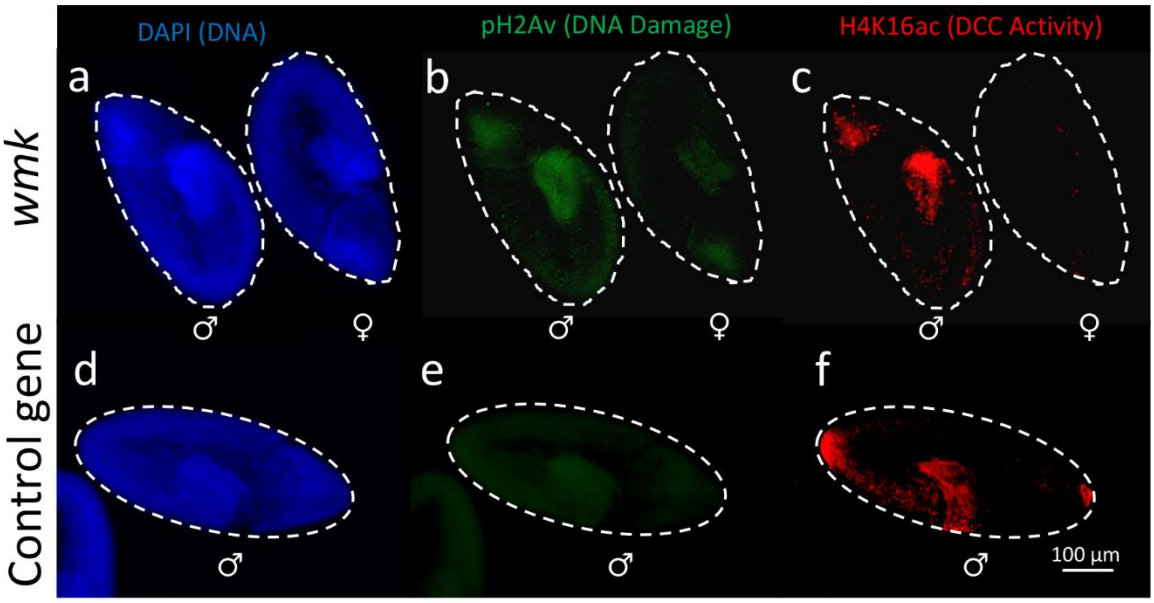
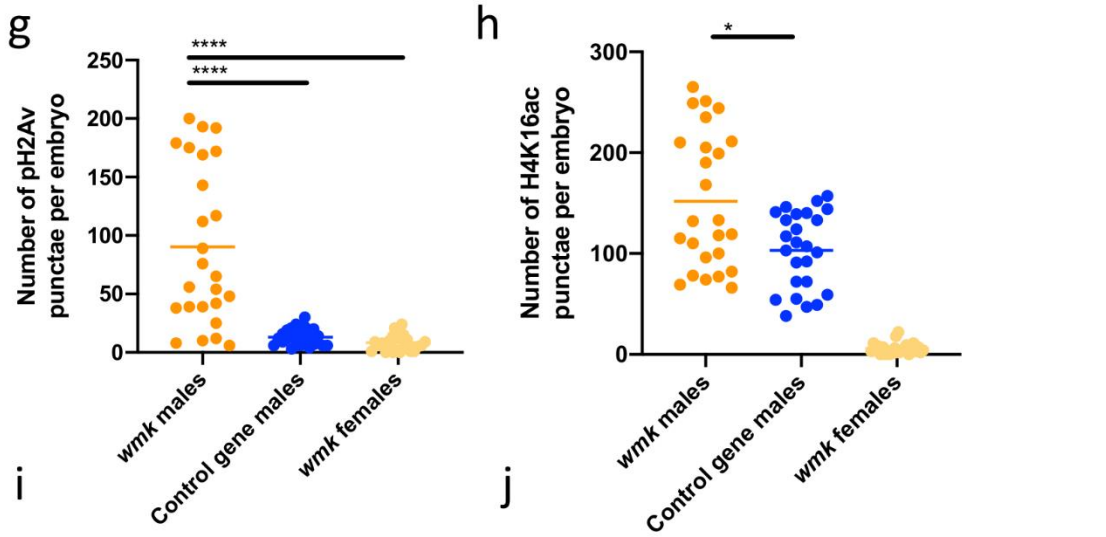


Figure 4-4. *wmk*-induced embryonic defects are enriched in males.
Data are from pooled embryos (both sexes, expressing and non-expressing, see methods) with either *wmk*, the control gene, or a WT background. (a) Graph quantitating the percentage of 3–4 h AED embryos (males or females) that have at least one defect (*wmk* males N = 228, control gene males N = 190, WT males N = 170, *wmk* females N = 240, control gene females N = 200, WT females N = 158). (b) Graph quantitating the sex ratio of viable embryos (not degraded, no visible defects) across two development times (1–2 h *wmk*, N = 105 m, 111 f; 1–2 h control gene, N = 30 m, 141 f; 1–2 h WT, N = 112 m, 115 f; 16–17 h *wmk*, N = 104 m, 154 f; 16–17 h control gene, N = 116 m, 120 f; 16017 h WT, N = 110 m, 108 f). m = male, f = female. Statistics were performed with a Chi-square test comparing the three genotypes within each category (male or female in (a) and 1–2 h or 16–17 h in (b)). These experiments were performed once. ** $p<0.01$, *** $p<0.001$, ****$p<0.0001$.

To further determine the similarity in lethality between the transgenic *wmk* and natural infection phenotypes, we assessed embryos for an association between DNA damage and dosage compensation. In previous work, male *D. bifasciata* embryos infected with *Wolbachia* exhibited an accumulation of DNA damage in association with dosage compensation[357]. We assessed *wmk*-expressing and control embryos 4-5 h AED for the same association (Fig. 4-5). Using the *armadillo* driver, we stained embryos with antibodies for pH2Av (phosphorylated histone H2Av, indicative of DNA damage) and H4K16ac (acetylation of histone H4 at lysine 16, primarily mediated on the X-chromosome by the male-specific dosage compensation complex or DCC). Males that express *wmk* have a greater number of pH2Av and H4K16ac punctae or foci than both *wmk*-expressing females and control gene-expressing males (Figs. 4-5Aa to 4-5h). The higher number of H4K16ac punctae may potentially reflect increased DCC activity in *wmk*-expressing embryos. An example set of images for a control gene female is shown in Fig. B-7. In addition, a significantly higher

proportion of the two types of punctae overlapped (Fig. 4-5i). This suggests a mechanism of death related to DNA damage that is associated with dosage compensation, as with natural infections. Within males, there is a cohort of *wmk* embryos that have a higher number of H4K16ac and pH2Av punctae (Figs. 4-5g and 4-5h). Interestingly, this proportion (~40%) is similar to the proportion of males that die according to adult sex ratios (Fig. B-3a). In addition, the H4K16ac and pH2Av punctae often overlapped with chromatin bridging, which is another phenotype previously observed in *D. bifasciata*[357]. The overlap happened more frequently in *wmk*-expressing males than females or control gene-expressing males (Fig. 4-5j). Taken together, results demonstrate that DNA damage is accumulating at sites of dosage compensation activity in *wmk*-expressing embryos.


DAPI (DNA)
pH2Av (DNA Damage)
H4K16ac (DCC Activity)
wmk
Control gene
a
b
c
d
e
f
♂
♀
100 µm


g
Number of pH2Av punctae per embryo
****
****
wmk males
Control gene males
wmk females
h
Number of H4K16ac punctae per embryo
*

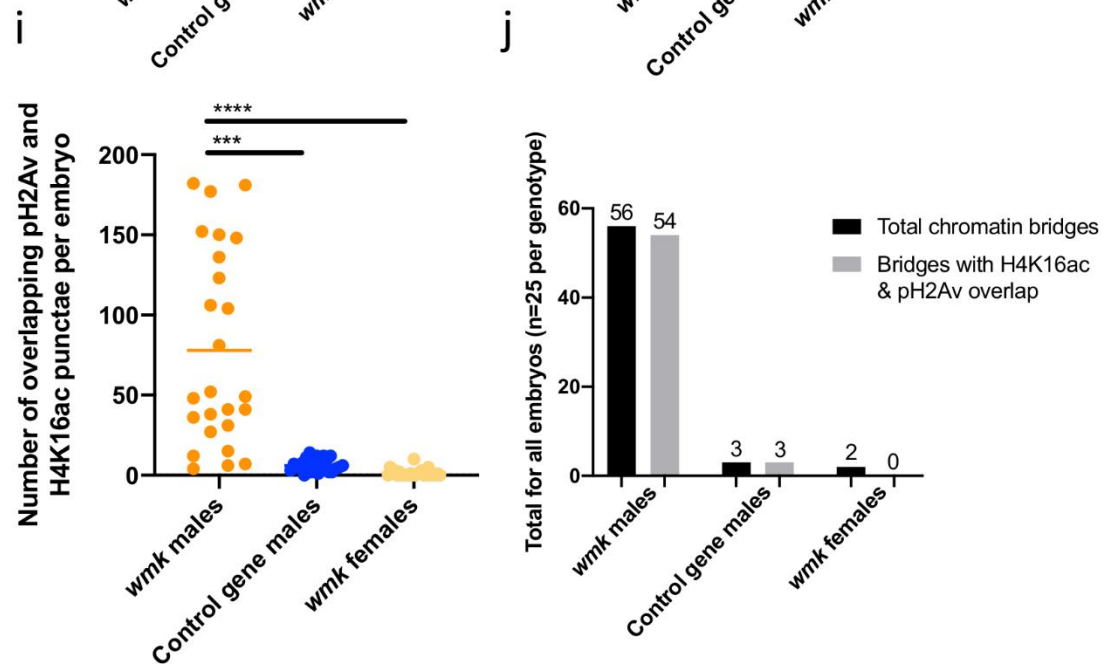

i
Number of overlapping pH2Av and H4K16ac punctae per embryo
****
***
j
Total for all embryos (n=25 per genotype)
56
54
3
3
2
0
Total chromatin bridges
Bridges with H4K16ac & pH2Av overlap

Figure 4-5. Transgenic expression of *wmk* causes DNA damage in association with H4K16ac.
Images and data are from embryos 4–5 h AED expressing a transgene under the *arm* driver. (a) DAPI DNA stain of male and female embryos, side-by-side, expressing *wmk*. Sexes determined by H4K16ac antibody. (b) pH2Av antibody staining of the same embryos as (a). The male has distinct punctae or foci, while the female does not. All embryos exhibit either a low level of autofluorescence at the same wavelength as the secondary antibody (Alexa 488) visible in both embryos or there is background staining. (c) H4K16ac antibody staining of the same embryos as (a). Distinct punctae are only visible in males, while females can exhibit low levels of staining. (d) DAPI DNA stain of control gene male. Sex determined by H4K16ac antibody. (e) pH2Av antibody staining of the same embryo as (d), with no distinct punctae and only autofluorescence or background staining visible. (f) H4K16ac antibody staining of the same embryos as (d). (g) Graph of the number of pH2Av punctae visible in each embryo. N = 25 embryos per genotype. Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction. (h) Graph of the number of H4K16ac punctae visible in each of the same embryos as measured in (g). Statistics are based on a Mann-Whitney U test comparing the two male categories. (i) Number of cases where pH2Av punctae directly overlapped with H4K16ac punctae in the same embryos as (g) and (h). Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction. (j) Graph of the total number of chromatin bridges and the total number of bridges with overlapping H4K16ac and pH2Av punctae in each of the three genotypes measured in (g-i). All images were taken at 20X zoom. This experiment has been performed once. *p<0.05, ***p<0.001, ****p<0.0001.

wmk *is expressed in* Drosophila *embryos infected with* Wolbachia

To establish a native expression profile for *wmk*, we measured relative transcription in *Wolbachia*-infected embryos fixed 4-5 h AED, which is the estimated time of death of most *wmk*-expressing male embryos. In *w*Mel-infected embryos, native *wmk* and control gene transcripts were approximately 10-fold lower than the highly expressed CI gene, *cif*A (Fig. 4-6a). There were no significant differences with either gene compared to the less abundant *cif*B gene transcript. Also, expression levels of the *wmk* and control transgenes are similar in uninfected *D. melanogaster*, and both are expressed significantly higher than native bacterial transcription of the same genes (Fig. 4-6b). Finally, *D. bifasciata* embryos infected with *w*Bif male-killing *Wolbachia* showed a *wmk*-like expression profile similar to *w*Mel, whereby the *cifA* homolog is expressed significantly higher than the *wmk* homolog (Fig. 4-6c). This suggests that differences in *cifA* vs *wmk* gene expression do not account for differences in reproductive parasitism phenotype where both CI and male killing can be induced by the same bacterial strain. Phenotypic differences may instead be determined by another factor such as host genotype.

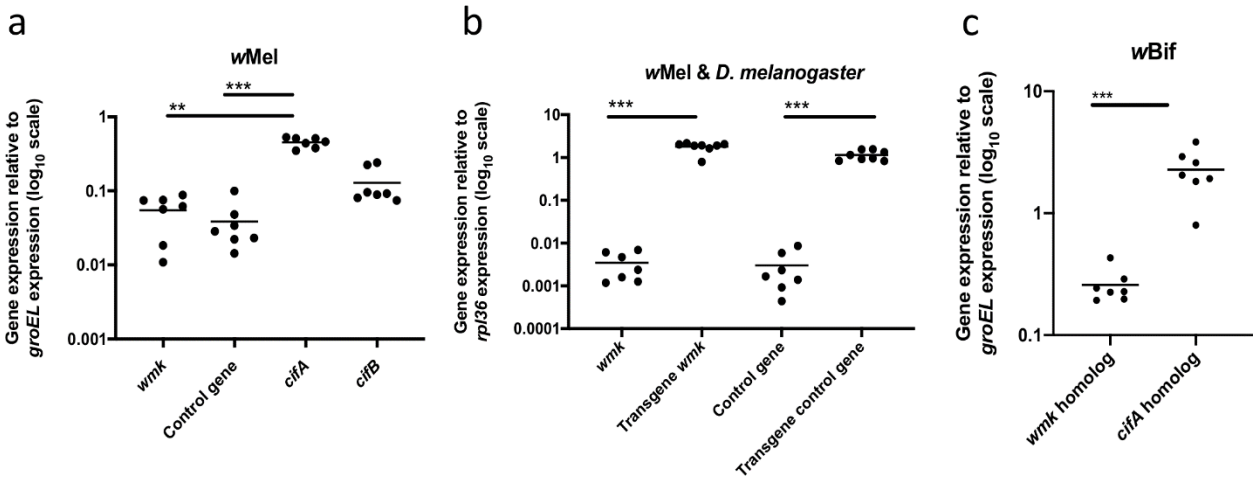


Figure 4-6. Native *Wolbachia* gene and transgene expression in embryos of *D. melanogaster* and *D. bifasciata*.

(a) Graph of native prophage WO and *Wolbachia* gene expression in *w*Mel-infected *D. melanogaster* embryos fixed 4–5 h AED (pooled male & female) compared to *Wolbachia gro*EL. Each point (n = 7) represents a pool of 30 embryos from a set of 10 mothers and 2 fathers. (b) Graph of (i) transgene expression in uninfected *D. melanogaster* embryos fixed 4–5 h AED versus (ii) native gene expression in samples from a, both compared to *Drosophila rpl*36 (pooled male, female, expressing, and non-expressing for transgenes). Each point (transgene n = 8, native n = 7) represents a pool of 30 embryos from a set of 10 mothers and 2 fathers. (c) Graph of *w*Bif *Wolbachia* gene expression in *D. bifasciata* embryos 4–5 h AED (pooled male & female) compared to *Wolbachia groEL*. Homologs to the control gene in this study and *cifB* were not measured as they are not present in the *w*Bif genome assembly. Each point (n = 7) represents a pool of 30 embryos from a set of 10 mothers and 2 fathers. Values denote $2^{-\Delta Ct}$. Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction. This experiment has been done once. **p<0.01, ***p<0.001.

The Wmk protein is a putative DNA-binding protein

Phyre2 protein modeling[440] predicts that Wmk from *w*Mel is globular and composed of α-helical secondary structures matching several transcriptional regulators, suppressors, and DNA-binding proteins (Fig. B-8a). The best match to known protein structures, based on both alignment confidence and sequence identity, is the *Salmonella* temperate phage Rep-Ant complex, a dimerized DNA- and peptide-binding repressor[441] (99.8% homology confidence, 19% sequence identity, Fig. B-8b). Wmk may function similarly as a bipartite protein where the dimers are physically connected, especially considering that single HTH domains typically dimerize and act as transcriptional regulators across domains of life[442]. Further, predicted structures of the Wmk homologs in *w*Bif (Fig. B-8c), *w*Inn/*w*Bor (same sequence, Fig. B-8d), and *w*Rec (Fig. B-8e) are all very similar to the structure from *w*Mel. Indeed, all exhibit a 5 α-helix bundle, connected by a

long, flexible linker to another 4 α-helix bundle. This is despite wide variation in amino acid sequence (e.g., *w*Bif Wmk has a 26.2% amino acid sequence identity to *w*Mel Wmk, which represent the most distantly related protein pair). Table B-6 shows amino acid pairwise percent identity between *w*Mel Wmk and homologs from known male-killers. This similarity in overall protein structure, despite sequence divergence, suggests that the homologs may retain the same general function with target(s) that are possibly divergent across host species, such as different DNA sequences of homologous genes. Wmk may also have another function that accounts for structural conservation despite sequence differences across divergent hosts.

To assess conservation in different regions of the protein, we also analyzed Wmk amino acid divergence across homologs, including that of *w*Bif and all homologs in Fig. B-1. There is relatively high sequence conservation overall across the protein (Fig. B-9a), but there are two areas of high variability adjacent to the two HTH DNA-binding domains that may be important for functional differences across strains or hosts (Fig. B-9b). In addition, although there is lower variation across DNA-binding regions relative to other parts of the protein, there is still variability that could account for differing abilities of homologs to cause a phenotype in one host versus another.

Discussion

This study reports twelve key results supporting *wmk* as a male-killing gene candidate: (i) *wmk* recurrently associates with genomic screens for reproductive parasitism; it is on the shortlists of candidate phage WO genes in *Wolbachia* male-killers and CI-inducers[158]. (ii) The *wmk* gene is found in all sequenced male-killers including the reduced phage WO genome of *w*Rec (which retains ~25% of the full phage WO genome) and the divergent phage WO genome of *w*Bif. (iii) *wmk* is common, divergent in sequence, and located in the eukaryotic association module of phage WO that is enriched with sequences predicted to contain eukaryotic function and homology[20]. In this region, *wmk* is a few genes away from the two causative cytoplasmic incompatibility genes, *cifA* and *cifB*, that modify arthropod gametes[158]. (iv) Transgenic expression of *wmk* consistently induces a sex-ratio bias, but the phenotype does not recapitulate other forms of reproductive parasitism. (v) No sex ratio bias results from expression of other transgenes tested thus far under the same expression system, making the phenotype specific to *wmk*. (vi) Canonical DNA defects are recapitulated under transgenic expression at the same time in development as natural systems.

(vii) *wmk* is naturally expressed in *w*Mel and *w*Bif embryos at the time the defects are known to occur in *D. bifasciata*. (viii) The Wmk protein is predicted to interact with DNA when DNA defects are a hallmark of *Wolbachia* male killing. (ix) *wmk* is unique to *Wolbachia*, and the *Wolbachia* male-killing mechanism has some unique phenotypic features compared to other male-killers. For example, the dosage compensation complex is not mislocalized in *Wolbachia* infection, but it is in *Spiroplasma* infection[357,358]. (x) The phenotype can be induced with drivers that yield approximately ten-fold variation in expression levels, indicating the highest *Act5c* levels of expression are not necessary for the phenotype. (xi) DNA damage is more common in *wmk* males than in controls and it is associated with H4K16ac, which parallels data in natural infections. (xii) Wmk's predicted structure is conserved across arthropod hosts despite sequence divergence, indicating it likely has conserved function.

Investigations into putative microbial male-killing genes have largely been hampered by an inability to culture or genetically manipulate intracellular bacteria and their mobile genetic elements. Recently, the gene Spaid in the endosymbiont *Spiroplasma poulsonii* was identified as a likely candidate underpinning killing of *D. melanogaster* males, possibly through misregulation of male dosage compensation[428]. Indeed, dosage compensation is an identified host target in *Spiroplasma* male killing[359,360], and may be involved in *Wolbachia* male killing as well, although likely through a different method such as increased activity rather than mislocalization that is typical of *Spiroplasma* infection[357]. It also appears that the *wmk*-mediated mechanism of male death may involve dosage compensation, as it recapitulates H4K16ac associations with DNA damage, but this remains to be confirmed with further experiments. Interestingly, *wmk* males have slightly more H4K16ac than their control gene counterparts, raising the possibility that death is correlated with either accelerated or a greater amount of H4K16ac. Whether this is true and whether the dosage compensation complex is directly or indirectly involved both remain to be determined.

Spaid is on a plasmid and has no homologs in *Wolbachia*, though it was previously noted that locus WD0633 in *w*Mel has similar protein domains consisting of ankyrin and OTU domains[428]. However, WD0633 was not predicted here to be on the shortlist of candidates for *Wolbachia* male-killing due to its absence in *w*Rec. *wmk* is also in the genome of a mobile element (phage WO), likely originated in *Wolbachia*, and has no homologs in *Spiroplasma*. This indicates that there could be an emerging trend of endosymbiotic reproductive parasitism genes and

candidates in mobile elements (including the *cifA* and *cifB* phage WO genes for CI). Both Spaid and *wmk* exhibit independent origins from each other. This finding is consistent with arguments that differences in observed male-killing phenotypes and sex determination systems of affected hosts may be due to distinct male-killing genes and/or mechanisms[258]. Other male-killing candidate genes may also exist. If so, they could support the observation that male-killing can independently arise in bacterial symbionts. Identification of additional genes and comparisons of their mechanisms is an important area of future work.

Wmk is also a putative DNA-binding transcriptional regulator (Fig. B-8), which is notable in light of previous studies demonstrating *Wolbachia*'s ability to modulate host transcription to induce various phenotypes. For example, *Spiroplasma*[359,360] and likely *Wolbachia*[357], kill males through the host dosage compensation complex, which is a critical mediator of transcriptional differences between male and female sex chromosomes. These reproductive parasites are therefore likely interfering with regulatory processes for host gene expression in males, which is a likely cause of male death. In addition, *Wolbachia* influences on host transcription have been implicated in the CI phenotype[443] and virus inhibition[146,444]. As *wmk* transgene expression similarly leads to DNA damage correlated with dosage compensation, it may follow a trend in the field of *Wolbachia* affecting the regulation or deregulation of host gene expression.

If *wmk* is the causative agent of male killing, then the *w*Mel genome could be multipotent and able to induce different phenotypes (e.g., CI and male killing) either in other hosts or under different environmental conditions. This premise remains to be evaluated in future studies. Assuming *wmk* is a bona fide male-killing gene, then some patterns about multipotency emerge. First and as noted earlier, *w*Mel and *w*Rec from *D. recens* are very closely related *Wolbachia* strains and have a 99.7% genome-wide identity[402]. Importantly, *w*Rec is a known multipotent strain that causes CI in its native host and male killing in a sister species[266]. While its genome has lost many prophage WO genes, it retains *wmk* and the *cif* genes that may underpin its multipotency, similar to *w*Mel. Second, while CI genes and phenotype often correlate, *wmk* is not always associated with male killing. *wmk* and its homologs are present in all sequenced male-killers, and they are also common in many other strains not known to cause male killing (Fig. 4-1, Fig. B-1a). In *w*Mel and potentially other strains, lack of male killing in native hosts is possibly due to host resistance to male killing, as is likely in *D. recens*[266]. Importantly, host suppression of male killing is common[266,328,329,345,430], presumably because of the evolutionary pressure on the host to develop

a counter-adaptation that avoids extinction. Therefore, though the *wmk* gene is more common than the male-killing phenotype, this would be expected if the frequency of resistance is indeed high. It is also possible that male killing is a multilocus trait that requires another gene to induce the phenotype in its natural context. Moreover, differences in *Wolbachia* titers, and/or insufficient expression of native *wmk* within *D. melanogaster* may contribute to the lack of male killing by *w*Mel, however this is unlikely given the similarly lowly-expressed *wmk* homolog in the *w*Bif male-killing strain. Finally, *wmk* and the *cif* genes are similarly disrupted, degraded, or lost in parthenogenesis-inducing *Wolbachia* strains *w*Uni from *Muscidifurax uniraptor* wasps, *w*Tpre from *Trichogramma pretiosum* wasps, and *w*Fol from *Folsomia candida* springtails. Therefore, multipotency is interestingly common for CI and male killing and will resultantly be rare in parthenogenesis strains.

There is considerable amino acid sequence divergence in Wmk homologs across several arthropod orders that harbor male-killing *Wolbachia*. One potential reason for the divergence is that if a single gene kills many or all of these hosts in nature, a premise which remains to be evaluated, it may be divergent due to selection to target the varied genetic and cellular bases of sex determination in these hosts. Second, if there is a single gene behind the phenotype, it could explain the relatively high frequency of host resistance since hosts would have to counter-adapt to one gene product rather than multiple products. Under antagonistic coevolution, *wmk* would evolve to kill males, the host adapts to resist the male killing, and *wmk* would follow suit and adapt again, continuing the evolutionary arms race. Third and in addition to coevolutionary bouts of *wmk* adaptation and host counter-adaptation, pleiotropy or multiple functions of *wmk* could also explain the sequence divergence in *wmk* homologs, especially in hosts that do not exhibit male killing.

Identification and further investigation of male-killing genes have relevance to translational applications in pest or vector control as male killing can theoretically be used in population suppression to crash target populations. Population modeling indicates that use of male killing in conjunction with other population-crashing techniques such as the Sterile Insect Technique (SIT), where sterilized males are released to compete with fertile males, could decrease the time to crash the population and increase the chances of success[373]. In this context, male killing genes might be used to transform an endosymbiotic microbe or host to either add or enhance male-killing ability. Alternatively, a male-killing infection could be established in a host where one does not natively exist. These techniques may be desirable in cases of invasive species of disease-

carrying mosquitoes or agricultural pests. Techniques like SIT can fail if males are not completely sterile or because of reduced mating competitiveness with fertile males[445,446]. Therefore, a two-pronged approach to simultaneously reduce viable matings in the wild (SIT) while killing off males (male-killing) could in principle be used to more effectively crash populations prone to SIT failure on their own[373], although this remains to be empirically evaluated.

There are many remaining questions for the future, including ones that are important for understanding a male-killing gene's role in host evolution and its potential in pest or vector control. First, is the *wmk* candidate gene in *Wolbachia* required for the phenotype in natural contexts? In the absence of the ability to knock out genes, it cannot yet be absolutely stated if *wmk* is used by bacteria to kill males in nature. Therefore, in addition to the transgenic expression, phenotype recapitulation, and sequence analyses demonstrated thus far, knocking out these genes in their resident genome will be important to assessing a change in phenotype. Second, can *wmk* homologs from related symbiont strains kill males? This will involve testing homologs in a genetically tractable host. Third, what is the exact mechanism of Wmk-induced male death? As *wmk* is annotated as a transcriptional regulator, it may act by controlling host transcription in a way that harms males. In addition, results indicate that the mechanism may involve dosage compensation. Fourth, what is the reason that transgenic *wmk* expression does not kill all males? Is it host resistance, inadequate expression patterns, divergence in host target or bacterial toxin gene sequence, or is another gene involved? We have tested a likely gene partner (WD0625) and multiple expression drivers (*Act5c*, *nanos*, *arm*, and MTD) to assess this, however no attempts so far have yielded answers. Finally, applications of male-killing bacteria or the genes to vector and pest control remain to be explored beyond population genetic theory[373].

The discovery of *wmk*-induced male death advances an understanding of the genes in the eukaryotic association module of prophage WO that interact with animal reproduction[158]. Moreover, male-specific lethality naturally occurs in many arthropods and has important influences on arthropod evolution[266,280,327,345,447,448], such as modifying mate choice and selecting for male resistance to the phenotype[176,258]. Male killing may also serve as a means to enhance population suppression methods for vectors or pests[373]. Thus, assessing male-killing gene candidates advances an understanding of the tritrophic crosstalk between phages, reproductive parasitic bacteria, and animals as well as their potential in arthropod control programs[373,426].

Materials and Methods

Experimental design

Most *Drosophila* experiments (unless otherwise noted) were set up with the following design. Crosses in each experiment were conducted by mating 10 female heterozygous *Act5c*-Gal4/CyO driver flies to 2 male homozygous transgene flies (both uninfected, unless otherwise noted; switching the gender for each genotype does not alter the effect). The offspring of these crosses were used for all experiments, except where noted. As the *Act5c*-Gal4/CyO driver strain is heterozygous, when driver flies are crossed to homozygous transgene flies, half of the offspring express the gene (those that inherit the *Act5c* driver gene that produces the Gal4 transcription factor), while the other half do not (those that inherit the CyO chromosome, which does not produce Gal4). Therefore, expressing males, expressing females, non-expressing males, and non-expressing females are expected in equal proportions under Mendelian inheritance. These four genotypes can only be visibly assessed in adulthood. Visually, embryos cannot be distinguished (except when fixed for microscopy with the Y chromosome FISH probe, when sex can be distinguished), while larvae can only be differentiated by sex.

Alongside several experiments, including the cytology in Figs. 4-3 and 4-4, sex ratios were measured concurrently. When flies were set up in the crosses described above, siblings were also set up in vials with CMY media. The protocol to measure sex ratios was then followed to obtain sex ratios side by side with these experiments. The results are in the extended data, where noted.

The maternal triple driver (MTD) was tested by crossing this homozygous driver strain to homozygous transgene flies in the same design as above. This crossing leads to transgene expression in all offspring because the driver is homozygous. Females expressing the transgene in their ovaries (MTD leads to targeted gene expression in the germline, specifically by loading embryos with the product) were then crossed to WT flies. Offspring were then quantified to measure sex ratios.

Comparative genomics and evolutionary analysis

Putative Wmk domains were identified by a CD-SEARCH of NCBI's Conserved Domain Database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). For the full-length analysis (Fig. B-1a), homologs were identified by a BLASTn of NCBI's nucleotide collection (nr/nt) and whole genome shotgun sequence (wgs) databases. The sequences reported were reciprocal best

BLAST hits with *w*Mel *wmk*. Partial sequences and/or those located at the end of a contig were excluded from downstream analysis. For the comparative genomic analysis, *wmk*, *cif*A, and *cif*B homologs were identified by manual annotations of prophage WO regions within known male-killing strains. Homology was confirmed by translating each gene and performing a BLASTP search against *w*Mel in NCBI. Only sequenced male-killing *Wolbachia* genomes in *Drosophila* were compared to demonstrate homologs clustering with gene synteny (Fig. B-1b). For both phylogenetic analyses, sequences were aligned using the MUSCLE plugin in Geneious Pro v8.1.7 and all indels were stripped. Trees were built using the MrBayes plugin in Geneious and were based on the best models of evolution, according to the corrected Akaike Information Criteria (AICc), as estimated by JModelTest and ProtTest v3.4.2, respectively. The models each predicted the GTR+I+G model for Fig. B-1a and the JTT+G model for Fig. B-1b, respectively. *w*Bif was excluded due to high sequence divergence. Protein modeling was performed with Phyre 2[440].

For the male-killer comparative genomics analysis, the entire *w*Bif draft assembly was searched for prophage WO-like regions. Five WO-like islands were found, and the genes in these regions were annotated using the NCBI BLASTP and conserved domain database. We then performed a 1:1 BLASTP of the annotated genes against query genomes. If it was present in the *w*Rec, *w*Inn, and *w*Bor genomes were searched for homologs, in the given order. If the gene was absent in one strain, it was marked as absent and excluded from further analysis. Genes were removed if they were: (i) absent in one or more of the strains (*w*Rec, *w*Inn, and *w*Bor), (ii) mobile elements (including IS elements, reverse transcriptases of group II intron origin, or recombinases), (iii) disrupted genes (frameshift with early stop codons) in one or more of the strains, and, (iv) if the E-value was less than E-20. See Table B-1 for a list of all removed genes along with rationale for exclusion.

Wolbachia *gene sequencing*

The *D. innubila Wolbachia* genome was sequenced from a single wild-caught female. Briefly, *D. innubila* were captured at the Southwest Research Station in Arizona over baits consisting of store-bought white button mushrooms (*Agaricus bisporus*). DNA was extracted using the Qiagen Gentra Puregene Tissue kit (#158689, Germantown, Maryland, USA). A genomic DNA library was constructed for several individuals using a modified version of the Nextera DNA Library Prep kit (#FC-121-1031, Illumina, Inc., San Diego, CA, USA) reagents[449]. DNA from an

infected female was sequenced on a fraction of an Illumina HiSeq 2500 System Rapid-Run to generate 14873460 paired-end 150 base-pair reads. Reads were aligned to a draft *D. innubila* genome and all non-aligned reads were assembled *de novo* using Spades[450]. Those contigs blasting to other *Wolbachia* accessions were retained as putative *Wolbachia* genomic contigs.

The *Wolbachia* genomes of *w*Bif and *w*Bor were sequenced from *D. bifasciata* (line bif-F-MK[451]) and *D. borealis* (line PG05.16[431]) respectively. Following the protocol developed in Ellegaard et al.[452], *Wolbachia* cells were purified from ~20 freshly laid (less than 2 hours) and bleach-dechorionated embryos by homogenizing them in phosphate-buffered saline solution (PBS) and conducting a series of centrifugation/filtration steps as explained in Ellegaard et al[452]. A multiple-displacement amplification was carried out directly on the bacterial pellet using the Repli-g midi kit (Qiagen). The amplified DNA was cleaned with QIAamp DNA mini kit (Qiagen). From each sample, both 3kb mate-pair and 50 bp paired-end DNA libraries were prepared and sequenced on a 454 Roche FLX (Department of Biochemistry, Cambridge, UK) and Illumina HiSeq2000 instruments (The Genome Analysis Center, Norwich, UK) respectively. The sequencing generated 203,565 and 239,485 454 mate-pair reads as well as 35,415,012 and 30,624,138 Illumina reads for *w*Bif and *w*Bor respectively. De novo hybrid assemblies combining 454 reads and a 10% subset of the Illumina reads were performed in Newbler (454 Life Sciences Corp., Roche, Branford, CT 06405, US). Contigs blasting to other *Wolbachia* accessions were retained as putative *Wolbachia* genomic contigs. Scaffolds were extended to fill regions with "N"s using GapFiller v.1-11[453].

The *Wolbachia* genome of *D. innubila* (*w*Inn) was sequenced by the R. Unckless lab. The *Wolbachia* genomes of *D. bifasciata* (*w*Bif) and *D. borealis* (*w*Bor) were sequenced by the F. Jiggins lab. The genomes will be published by the respective contributors at a later date, and only the phage WO gene regions involved in this publication are publicly available (the regions in Fig. 4-1).

Drosophila *strains*

The *Wolbachia* transgene strains were generated as described previously [158]. WD0626 (*wmk*) and WD0034 (control gene) were both inserted into an attP site in the BSC8622 (WT) line of genotype y^1w^{67c23}; P[CaryP]P2 obtained from the Bloomington Drosophila Stock Center. WD0625 was inserted into the BSC9723 strain, with a genotype of y^1M[vas-int.Dm]ZH-2A w*;

PBac[y+-attP-3B]VK00002. WD0508 was inserted into the y^1M[vas-int.Dm]ZH-2A w*; P[CaryP]attP40 line. The genes were inserted into various strains to facilitate creation of strains that contain more than one gene homozygously. The *Act5c*-Gal4/CyO driver line is the same background as BSC3953, which is $y^1 w$*; P[Act5C-GAL4-w]E1/CyO. The maternal triple driver (MTD) strain BSC31777, genotype P[w[+mC]=otu-GAL4::VP16.R]1, *w*[*];P[w[+mC]=GAL4-nos.NGT]40; P[w[+mC]=GAL4::VP16-nos.UTR]CG6325[MVD1], was provided by J. Nordman. The *nanos*-Gal4 strain used in S4B and S4C Fig was previously described [158]. The *arm*-Gal4 driver strain BSC1560 is w[*]; p[w[+mW.hs]=GAL4-arm.S]11. The infected *D. bifasciata* flies were provided by G. Hurst and are infected with male-killing *Wolbachia*. The male-killing flies are maintained with males from a concurrently reared uninfected line also provided by G. Hurst.

Drosophila *rearing*

D. melanogaster were reared on 4% cornmeal (w/v), 9% molasses (w/v), 1.6% yeast (w/v) (CMY) media. The flies developed at 25°C at 80% humidity with a 12 h light/dark cycle. Virgin flies were stored at room temperature after collections. During virgin collections, stocks were maintained at 25°C during the day and at 18°C at night. *Wolbachia*-uninfected transgene or driver lines were generated via tetracycline treatment of infected lines as described previously[158]. *D. bifasciata* are maintained on CMY media at room temperature.

Sex ratio measurements

To assess the ability of the gene candidates to alter sex ratios, twenty replicates of 10 uninfected, 4-7 day old female driver flies and 2 uninfected, 1-2 day old male transgene flies were set up in vials with CMY media. They were left on the media to lay eggs for 36 h at 25°C, at which point adults were discarded. Once the offspring emerged, they were scored for both sex and expression or non-expression (if applicable), which was determined by presence or absence of the CyO wing phenotype as well as with eye color markers associated with *Act5c*-Gal4 and the transgene insertion. Any vials with fewer than 50 adult offspring were removed from the analysis, as this indicates either poor egg laying or abnormally low egg hatching (average = 120 offspring).

Hatch rates

Extended data hatch rates (Figs. B-4b and B-4c) were performed as previously described with the *nanos*-Gal4 driver[158]. The *nanos* driver was used to test induction of CI instead of *Act5c*-Gal4/CyO because it is expressed more specifically in the gonads where CI is induced[158].

Embryo cytology

For Figs 4-3 and 4-4, eight stock bottles were set up per genotype, each with 60 uninfected, 4-7 day old *Act5c*-Gal4/CyO females and 12 uninfected, 1-2 day old transgene or WT males. Grape juice agar plates, made as described previously[158], with a small amount of baker's yeast (Red Star) placed on each bottle opening and fixed with tape. They were then placed with the grape plate down in a 25°C incubator overnight (~16 hr). The grape plates were then replaced with fresh plates and fresh yeast. The flies were then allowed to lay eggs in 1 h increments, replacing the previous plates with fresh ones each time. They were then allowed to sit at room temperature for 1 h (embryos 1-2 h old), 3 h (3-4 h old), or 16 h (16-17 h old). Once they had reached the desired point in development, the embryos were fixed and stained, using a slight modification of the protocol outlined by Cheng et al. 2016[358]. Briefly, the embryos were dechorionated in 50% bleach and fixed for 15 minutes in a 1:1 4% paraformaldehyde:heptane mixture while shaking on a tabletop vortexer at about 150 rpm. The solution was discarded, and the embryos were then devitellinized in a 1:1 heptane:methanol mixture by shaking vigorously for one minute. The solution was removed, and the embryos were placed in fresh methanol and stored at 4°C until the next steps were done, at least 16 h later. Then, the methanol was removed and the embryos were rehydrated in a series of methanol:water solutions, in the order of 9:1, then 1:1, then 1:9, each for 15 minutes while mixing on a Nutator. They were then treated with 10 mg/mL RNase A (Clontech Labs) by incubating them at 37°C for 2-3 hr with enough RNase solution to cover the embryos. Once the RNase was removed, the embryos were washed three times for 5 min each in PBST (1X PBS, 0.1% Tween 20), while mixing on the Nutator. They were then re-fixed in 4% paraformaldehyde for 45 minutes with mixing and were then washed or incubated with several solutions with mixing on the nutator. First, they were washed three times in saline-sodium citrate/Tween 20 buffer (SSCT, 2X SSC buffer, 0.1% Tween 20) for 10 minutes each. They were then incubated with a series of SSCT/formamide solutions for 10 minutes each in the following order: 80% SSCT/ 20% formamide, 60% SSCT/ 40% formamide, 50% SSCT/ 50% formamide. Then fresh 50% SSCT/

50% formamide was added and the embryos were incubated at 37°C for 1 h. The solution was removed, and the embryos were then hybridized with the Y-chromosome FISH probe. This was done by mixing 36 µL FISH hybridization solution (1g dextran sulfate, 1.5 mL 20X SSC, 5 mL formamide, to 15 mL with DNase-free water)[454], 3 µL DNase-free water, and 1 µL 200 ng/µL Y-chromosome FISH probe (sequence 5'-AATACAATACAATACAATACAATACAATAC-3' synthesized with Cy5 conjugated to the 5'end (IDT)) using the sequence published by Cheng et al. 2016[358]. Hybridization was done in a thermocycler by denaturing at 92°C for 3 min, followed by hybridizing at 37°C overnight (~16 h). Then, the embryos were again washed in a series of solutions on the nutator. They were done in the order of three 15 min 50% SSCT / 50% formamide washes, one 10 min 60% SSCT / 40% formamide wash, one 10 min 80% SSCT / 20% formamide wash, and three 10 min SSCT washes. They were then mounted on glass slides with ProLong Diamond Antifade (Life Technologies, P36970) mounting media that contained 1 µg/mL propidium iodide (Sigma Aldrich).

Imaging was performed at the Vanderbilt University Cell Imaging Shared Resource (CISR) with a Zeiss LSM 510 META inverted confocal microscope. Images are of a single plane. Image analysis and preparation was done with ImageJ software. Image brightness and contrast were adjusted for visibility, but adjustments were applied equally across each whole image.

For Fig. 4-5, a different fixing and staining protocol was used. Eight bottles were set up per genotype with 60 uninfected *armadillo(arm)*-Gal4 females crossed to 12 uninfected *wmk* or control gene males with a small amount of baker's yeast (Red Star) placed on each bottle opening and fixed with tape. They were then placed with the grape plate down in a 25°C incubator overnight (~16 hr). The grape plates were then replaced with fresh plates and fresh yeast. The flies were then allowed to lay eggs in 1 h increments, replacing the previous plates with fresh ones each time. They were all aged to 4-5 h AED. Once they had aged to the desired point in development, they were fixed and stained using the protocol described in Hall & Ward[455]. Embryos were dechorionated for 2 min in 50% bleach and rinsed with water. They were then fixed with shaking in 1:1 4% paraformaldehyde to heptane at room temperature for 20 min. The bottom paraformaldehyde phase was removed and methanol was added in equal volume to the remaining heptane and embryos. They were then devitellinized by shaking vigorously for 20 s. Embryos were stored in methanol at 4°C until staining. Staining was performed by first removing the methanol and rinsing with 750 µL blocking solution (Vector Laboratories Animal-Free blocking solution

SP5030). The embryos were then rinsed in 1X PBS twice. The PBS was removed and the embryos were permeabilized in 750 µL blocking solution for 30 min at room temperature with rocking. The blocking solution was removed and the embryos were rinsed with 1X PBS once. The embryos were then incubated with primary antibodies in 500 µL blocking solution overnight at 4°C with rocking. The antibodies included histone H2AvD pS137 antibody (1:100, Rockland 600-401-914), anti-acetyl-histone H4 (Lys16) antibody or H4K16ac (1:100, Millipore Sigma 07-329), and Sxl antibody (1:20, DSHB M18). The Sxl antibody developed by P. Schedl was obtained from the Developmental Studies Hybridoma Bank, created by the NICHD of the NIH and maintained at The University of Iowa, Department of Biology, Iowa City, IA 52242. In cases where primary antibodies were raised in the same animal, sequential staining was performed. After overnight staining with one antibody, the steps were repeated beginning with the initial blocking step for the second antibody.

After overnight staining, the embryos were washed in 1X PBS three times at room temperature with rocking for 5 min each. They were then incubated with 750 µL blocking solution for 30 min at room temperature with rocking. The blocking solution was removed and the embryos were rinsed in 1X PBS once. The embryos were then incubated with secondary antibodies in 500 µL blocking solution at room temperature with rocking for 1 h out of the light (all subsequent steps are also out of the light). The antibodies included goat anti-mouse IgG with Alexa Fluor 647 (1:500, abcam ab150115), goat anti-rabbit IgG with Alexa Fluor 594 (1:500, Invitrogen A11037), and goat anti-rabbit IgG with Alexa Fluor 488 (1:500, Invitrogen A11034). The embryos were then washed three times with 1X PBS at room temperature with rocking for 5 min each. They were then incubated with 750 µL blocking solution for 30 min at room temperature with rocking. The embryos were then rinsed once in 1X PBS. The embryos were then stained with 1µg/mL DAPI (Invitrogen D1306) for 10 min with rocking at room temperature. Embryos were then washed three times in 1X PBS for 10 min each with rocking at room temperature. They were then mounted on glass slides with ProLong Diamond Antifade (Life Technologies, P36970) mounting media.

Imaging was performed using a Keyence BZ-X710 Fluorescence Microscope and all images are a single plane. Images were taken at 20X magnification. Quantification of punctae was done by manually focusing on several planes that encompassed all punctae and quantifying punctae with overlapping signals. Images were analyzed using Keyence analysis software. Image

brightness and contrast were adjusted and dehazing software was used for visibility, but adjustments were applied equally across each whole image.

Gene expression

Gene expression in embryos from Fig. 4-6 was measured in each of four groups. Group 1 was generated in crosses between *Act5c*-Gal4/CyO uninfected females crossed to *wmk* uninfected males. Group 2 was generated in crosses between *Act5c*-Gal4/CyO uninfected females crossed to control gene uninfected males. Group 3 was generated by crossing y^1w^* infected females to y^1w^* uninfected males. Group 4 was generated by crossing *w*Bif-infected *D. bifasciata* females to uninfected *D. bifasciata* males. Gene expression for Fig. B-3b was set up using two groups with either *Arm*-Gal4 or *Act5c*-Gal4/CyO uninfected females crossed to *wmk* males. For each group, 8 bottles were set up with 10 females and 2 males. A grape juice agar plate[158] with yeast was placed in each bottle. These were placed in a 25°C incubator overnight (16 h) for *D. melanogaster* or kept at room temperature (23°C) for *D. bifasciata*. 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 or 18°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 the same bottle) 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 B-4. 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. For each gene measured, a standard curve was produced with known concentrations alongside samples with unknown concentrations. Primers are listed in Table B-3. Differences in gene expression were done by calculating $2^{-\Delta ct}$ (difference in ct values of two genes of interest).

Confirmation of gene expression in adults from Figs. B-2c and B-2e was done similarly. Samples were obtained by flash freezing adult offspring laid by siblings of the flies used in Fig. B-2a. Samples from Fig. B-2b were from pooled, whole-body extractions from three males of each genotype. Samples from Fig. B-2c were from pooled, whole-body extractions from three females of each genotype. Samples from Fig. B-2e were from pooled, dissected ovaries of six adult female siblings of flies of flies used in Fig. B-2e for each genotype. Samples were flash frozen in liquid

nitrogen and then was processed (RNA extraction, DNase treatment, and cDNA treatment) as above. PCR was performed against positive controls (extracted DNA), negative controls (water), RNA, and cDNA. Gel image brightness and contrast were adjusted for visual clarity, but adjustments were applied equally across each whole image.

Protein conservation

Protein conservation was calculated with the Protein Residue Conservation Prediction Tool[456]. Amino acid sequences from Fig. B-1 along with the *w*Bif Wmk homolog sequence were aligned using a MUSCLE alignment in Geneious Prime version 2019.1. This alignment was uploaded to the prediction tool with the following settings: Shannon entropy scores, a window size of zero, and no sequence weighting. Conservation values were then input into GraphPad Prism version 8 for visualization. HTH regions were indicated using the amino acids predicted to be in the domains according to the NCBI annotation of *w*Mel Wmk.

Statistical analysis

Statistical analyses were done using GraphPad Prism software (version 5 or 8) or GraphPad online tools, unless otherwise noted. For comparisons among only two data categories, we used the two-tailed, non-parametric Mann-Whitney U test. For comparisons with more groups, a non-parametric Kruskal-Wallis one-way analysis of variance was used, followed by Dunn's test for multiple comparisons, if significant. In cases of comparisons among groups where only a single measurement was taken per group (such as cytology experiments), a Chi-square test was used. Exact tests used and other important information is listed in the figure legends of each experiment.

Data availability

Accession numbers at NCBI are listed in supplemental information.

Acknowledgements

We would like to thank J. Nordman for providing MTD flies. We also thank J. On for assistance in collecting flies and various colleagues in the research community and other members of the Bordenstein lab for comments and critiques. Additionally, we thank G. Hurst for providing *D. bifasciata* flies and the anonymous reviewers for their helpful feedback.

Chapter 5

Transgenic testing does not support a role for additional candidate genes in male killing or cytoplasmic incompatibility

Contributing Authors: Jessamyn I. Perlmutter, Jane E. Meyers, and Seth R. Bordenstein

Abstract

Endosymbiotic bacteria in the genus *Wolbachia* remarkably infect nearly half of all arthropod species. They spread in part because of manipulations of host sexual reproduction that enhance the maternal transmission of the bacteria, including male killing (death of infected males) and unidirectional cytoplasmic incompatibility (CI, death of offspring from infected fathers and uninfected mothers). Recent discoveries identified several genes in prophage WO of *Wolbachia* (*wmk*, *cifA*, and *cifB*) that fully or partially recapitulate male killing or CI when transgenically expressed in *Drosophila melanogaster*. However, it is not yet fully resolved if other gene candidates contribute to these phenotypes. Here, we transgenically test ten additional gene candidates for their involvement in male killing and/or CI. Results show that despite sequence and protein architecture similarities or comparative associations with reproductive parasitism, transgenic expression of the candidates does not recapitulate male killing or CI. Sequence analysis across Wmk and its closest relatives reveals amino acids that may be important to its function. In addition, evidence is presented to propose new hypotheses regarding the relationship between transcript length of *wmk* and its ability to kill a given host, as well as copy number of *wmk* homologs within a bacterial strain that may be predictive of host resistance. Together, these analyses continue to build the evidence for *wmk*, *cifA*, and *cifB* as the major genes that thus far cause reproductive parasitism in *Wolbachia*, and the transgenic resources provide a basis for further functional study of phage WO genes.

Importance

Wolbachia are widespread bacterial endosymbionts that manipulate the reproduction of diverse arthropods to spread through a population and can substantially shape host evolution. Recently, reports named three prophage WO genes (*wmk*, *cifA*, and *cifB*) that transgenically recapitulate many aspects of reproductive manipulation in *Drosophila melanogaster*. Here, we transgenically test ten additional gene candidates for CI and/or male killing in flies. Results yield

no evidence for the involvement of these gene candidates in reproductive parasitism, bolstering the evidence for *cif* and *wmk* genes as the major factors involved in their phenotypes. In addition, evidence supports new hypotheses for prediction of male killing phenotypes or lack thereof based on *wmk* transcript length and copy number. These experiments inform efforts to understand the full basis of reproductive parasitism for basic and applied purposes and lay the foundation for future work on the function of an interesting group of *Wolbachia* and phage WO genes.

Introduction

Some of the most widespread microbial symbioses on the planet occur between invertebrates and various microbes that manipulate host reproduction[23]. These reproductive parasites hijack host cellular processes and alter host reproduction to facilitate their spread. They include a variety of maternally-inherited bacterial, fungal, and viral endosymbionts that infect a large number of arthropod hosts including all major groups of insects and arachnids[457]. Of these microbes, the most common are the genus *Wolbachia*, which are obligate intracellular bacteria that manipulate host reproduction in a variety of ways[397,458]. There are at least four main phenotypes, including, (i) male killing (selective killing of male hosts), (ii) feminization (physical development and reproduction of genetic males as females), (iii) parthenogenesis (asexual reproduction of females), and (iv) unidirectional cytoplasmic incompatibility (CI, death of offspring when infected males mate with uninfected females or infected females harboring incompatible strains, and "rescue" from death in matings between parents infected with compatible strains). Each phenotype facilitates spread of the bacteria by either increasing the fitness of infected females through induction of a female-biased sex ratio (i-iii) or decreasing the fitness of uninfected females through a reduction in viable offspring (iv).

Of these phenotypes, two in particular have current or potential use in arthropod pest and vector control efforts and are important to the basic biology of both host and microbe, making them the subject of diverse research interest. CI, the most widespread phenotype, is currently deployed in *Aedes albopictus* and *A. aegypti* mosquitoes to reduce the incidence of vector-borne diseases[183,459-461]. These efforts have achieved early success[183,371,461], but there are potential challenges for widespread applications of CI-based vector control including potential difficulty in spreading CI-*Wolbachia* strains to recalcitrant host species[6,266,341]. In addition, male killing is a potential adjunctive or standalone control method. Although it has not yet been tested empirically

in arthropods, population modeling suggests male killing could be especially useful in a two-pronged approach alongside CI or sterile insect technique (SIT)[373]. CI and male killing also have important consequences for host evolution and ecology. As CI can kill offspring from crosses between infected males and uninfected females or between infected males and females harboring incompatible strains of *Wolbachia*, CI can be a barrier to gene flow between populations or incipient species[175,181]. Male killing, on the other hand, may lead to evolutionary outcomes including host extinction, loss of the male killer[327], and host development of heritable resistance to male killing[266,329,345]. In addition, female-biased populations may exhibit altered sexual selection. For example, infected *Hypolimnas bolina* female butterflies become promiscuous, can form lekking swarms, and display mate-attracting behaviors[174,176].

Given the aforementioned relevance to both applied and basic research, there is considerable interest in the genetics of reproductive parasitism[248] and phage WO in particular[20,158-161,402,462,463]. Phage WO has a unique genome compared to other phages because it includes eukaryotic association module (EAM) that is enriched with genes annotated or demonstrated to have eukaryotic function or homology[20]. Many of the EAM genes are unique to this bacteriophage and putatively encode functions that underlie host-symbiont interactions. Indeed, the genes underlying the CI phenotype (*cifA* and *cifB*, *cytoplasmic incompatibility factors A* and *B*, loci WD0631 and WD0632), are just a few genes away from the male-killing gene candidate (*wmk*, *WO-mediated killing*, locus WD0626) in the EAM region of prophage WO[158,160,161]. The two *cif* genes synthetically recapitulate the full CI and rescue phenotypes when expressed transgenically in *Drosophila melanogaster*[158,161]. Similarly, transgenic *wmk* expression specifically and consistently kills a third of male hosts and preferentially induces cytological defects in male embryos that are typical of natural infection[160]. In addition, a *Spiroplasma* male-killing gene, *Sp*AID, was recently reported on a plasmid and likely functions via interference with host dosage compensation[248]. Significantly, all of these genes are unique in nature, with specialized functions and no known homologs in other organisms, and thus they represent new frontiers in understanding host-endosymbiont biology.

Importantly, although the genes thus far recapitulate several cytological, biochemical, and embryonic phenotypes of natural infection, the genetic basis of reproductive parasitism may not be fully resolved. Additional gene candidates from comparative genomic analyses exist, and modifier genes in phage WO or *Wolbachia* may alter the penetrance of the phenotypes. For

example, the *wmk* gene is a candidate for *Wolbachia*-induced male killing due to its recapitulation of many aspects of the natural phenotype, including male-biased embryonic defects and lethality, and associations between dosage compensation activity and DNA damage. However, *wmk* expression kills over a third of gene-expressing males instead of all males under the conditions tested thus far[160]. The incomplete penetrance could be due to inadequate transgenic expression levels or patterns, host resistance, or involvement of another gene in the phenotype[160]. Previous work tested different *wmk* expression levels, but results showed that increased expression levels lead to a similar phenotype[160]. Notably, there are many connections between the CI and male-killing cytological defects (such as chromatin bridging)[158,160,356,392] that suggest they may have overlapping functions, but the basis of these connections remain unclear. In addition, many strains of *Wolbachia* are multipotent and thus induce either male killing or CI depending on the host or environment[262,266,329,341]. Further, a previous comparative genomic analysis for CI-associated genes demonstrated that *wmk* was shared across CI-causing genomes[158], and *cifA* was identified as a top candidate in a comparative genomic analysis for the genes underpinning male killing[160]. Not all additional candidates have been tested, nor for all putative phenotypes, and thus, there may be other genes that recapitulate CI or male killing. The nature of any putative relationship between CI and male killing genes is also unclear.

Several genes were previously identified that are moderately associated with reproductive parasitism in *Wolbachia*, but they were not empirically tested for function. These candidates were identified through similarity to *Sp*AID or the CI proteins[248], homology to the *wmk* gene, or identification in previous comparative genomic analyses for genes associated with male killing[160]. Here, we analyze and transgenically test these gene candidates for recapitulation of reproductive parasitism to assess the hypothesis that phage WO contains additional genes that mediate parasitism of host reproduction.

Results

Many additional w*Mel genes are candidates for reproductive parasitism*

Although several studies have recently identified genes that recapitulate reproductive parasitism phenotypes, additional male-killing candidates have been previously reported or are reported here, albeit with lower support for a genotype-phenotype association (Table 5-1). The reasons for inclusion of these genes can be broken down into several categories: (i) predicted

protein similarity to the *Sp*AID male-killing toxin, probability of type IV secretion, and presence of an OTU deubiquitinase domain similar to the CI genes (WD0633); (ii) additional homologs of *wmk* within the *w*Mel *Wolbachia* genome (WD0622, WD0623, WD0255); (iii) candidates identified through a previous male-killing comparative genomic analysis (WD1243, WD0296, WD0550, WD0631 (*cifA*), WD0628, WD0627); and (iv) *wmk* with a putative alternative start codon (identified and described here). Of the loci identified, two are in the WOMelA prophage region, seven are in the WOMelB prophage region that includes the EAM, and two are in the *Wolbachia* chromosome of the *w*Mel strain, but are in the prophage region of the *w*Bif male-killing *Wolbachia* strain of *D. bifasciata* (Fig. 5-1).

Table 5-1. List of gene candidates for male killing.
List of gene candidates for *Wolbachia* / phage WO male killing, their putative functions or domains, the basis for their inclusion, and any publications or figures in which they were identified as candidates. CI: cytoplasmic incompatibility. MK: male killing.

Gene	Putative Function(s) or Domain(s)	Reason(s) for Inclusion	Identifying Publication or Figure
WD0626 (*wmk*)	HTH DNA-binding TF	Previous transgenic testing	160
WD0633	OTU, ankyrin repeats	Similar domains to *Sp*AID	248
		Previous correlation with CI	464
		Similar to known T4SS effectors	Fig S2
WD0622	HTH DNA-binding TF	Homolog of *wmk*	160
WD0623	HTH DNA-binding TF	Homolog of *wmk*	160
WD0255	HTH DNA-binding TF	Homolog of *wmk*	160
WD1243	Putative phospholipase D or nuclease	Genomic analysis for MK candidates	160
WD0296	Recombination-promoting nuclease (Rpn)	Genomic analysis for MK candidates	160
WD0550	Ankyrin repeats	Genomic analysis for MK candidates	160
WD0628	Hypothetical protein	Genomic analysis for MK candidates	160
WD0627	Recombination-promoting nuclease (Rpn)	Genomic analysis for MK candidates	160
WD0626 (*wmk*), alternative start codon	HTH DNA-binding TF	Previous transgenic testing	Fig 6, S3
WD0631 (*cifA*)	DUF, catalase-rel, STE TF	Similar phylogeny	Fig 5

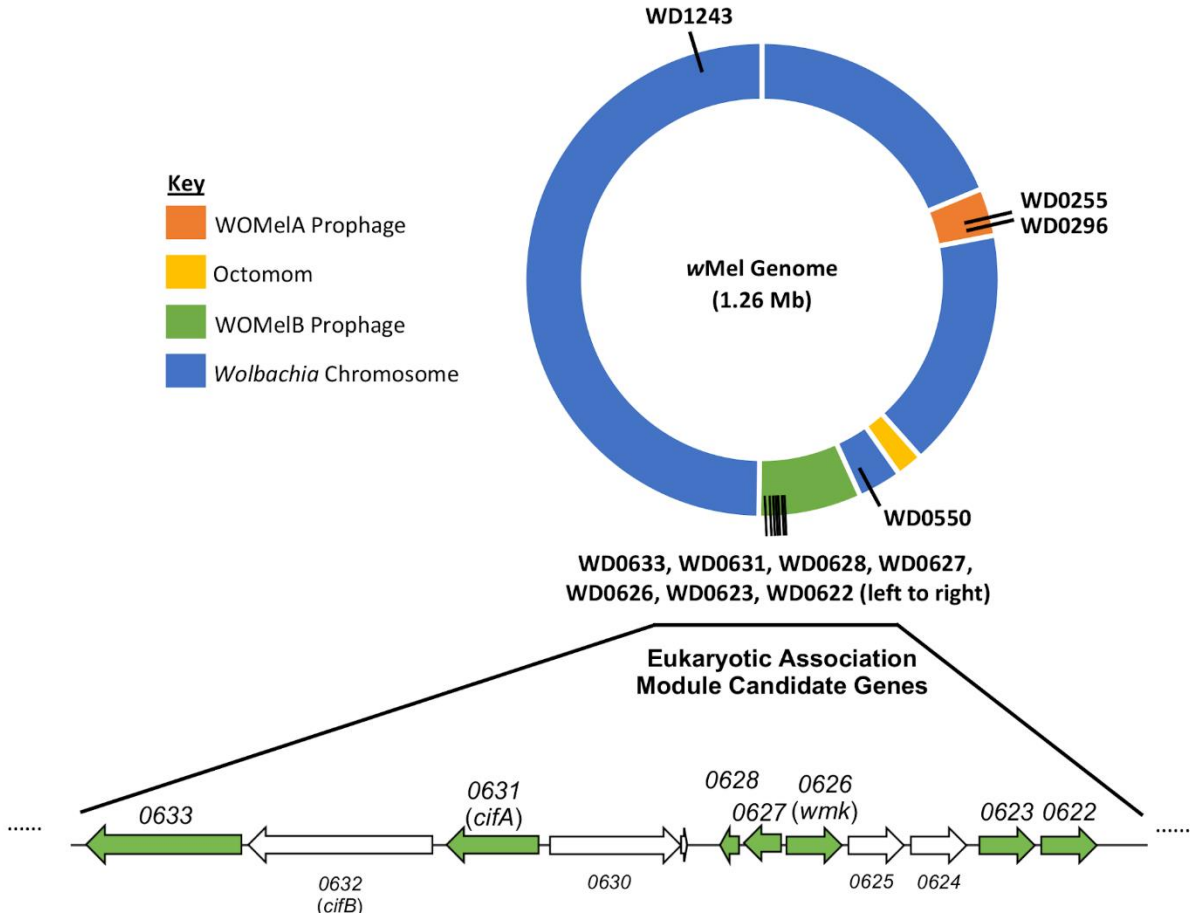


Figure 5-1. Map of gene candidates assessed for reproductive parasitism across the *w*Mel genome.
Prophage WO regions are shown in their indicated colors. Gene positions, indicated by black lines, are approximate. The white line at the top indicates the first nucleotide position in the genome. The WOMelB prophage region is expanded below to show the relative positions of these genes. Genes are roughly to scale, with candidates in green and non-candidate genes in white. Different arrow directions indicate location on opposite DNA strands.

The Sp*AID homolog WD0633 in prophage WO does not transgenically recapitulate male killing or CI*

The candidate gene WD0633 was identified in a previous publication that reported *Sp*AID as a strong candidate for the *Spiroplasma* male-killing toxin[248]. Although the authors did not find any homologs in *Wolbachia* based on the full gene sequence, they noted that WD0633 shares putative protein domain features such as an OTU deubiquitinase domain and several ankyrin repeats. Despite this similarity in putative domain identities, the overall protein architecture is

different due to the presence of fewer putative ankyrin repeats and different localization of the deubiquitinase domain in WD0633 (Fig. 5-2a). In addition, Blastp results of *Sp*AID do not show any full homologs to the gene in organisms other than *Spiroplasma*. However, genes on mobile elements such as phages or plasmids (*Sp*AID is reportedly on a plasmid) are often developed by fusion of gene sequences from several different sources[20]. Therefore, we performed Blastp searches on different regions of the protein. Results showed that the OTU domain had weak homology to *Wolbachia* proteins, while other regions only had homology to *Spiroplasma* proteins. An unrooted Bayesian tree demonstrates that homologs cluster by bacterial genus (Fig. C-1a). The homology suggests that there may have been gene exchanges between these two genera, although the direction of any putative gene exchange is undetermined. The likelihood of a gene transfer event is not unreasonable given that both bacteria can infect the same host organisms[289,465]. In contrast, the WD0633 full protein and OTU domain have no significant homology to *Spiroplasma* proteins, leaving no indication of a relationship with *Spiroplasma*. Thus, there is a potential link between *Sp*AID and other *Wolbachia* protein sequences, but the results support previous findings that WD0633 and *Sp*AID are not true homologs[248], nor is WD0633 a homolog of any other known *Spiroplasma* protein.

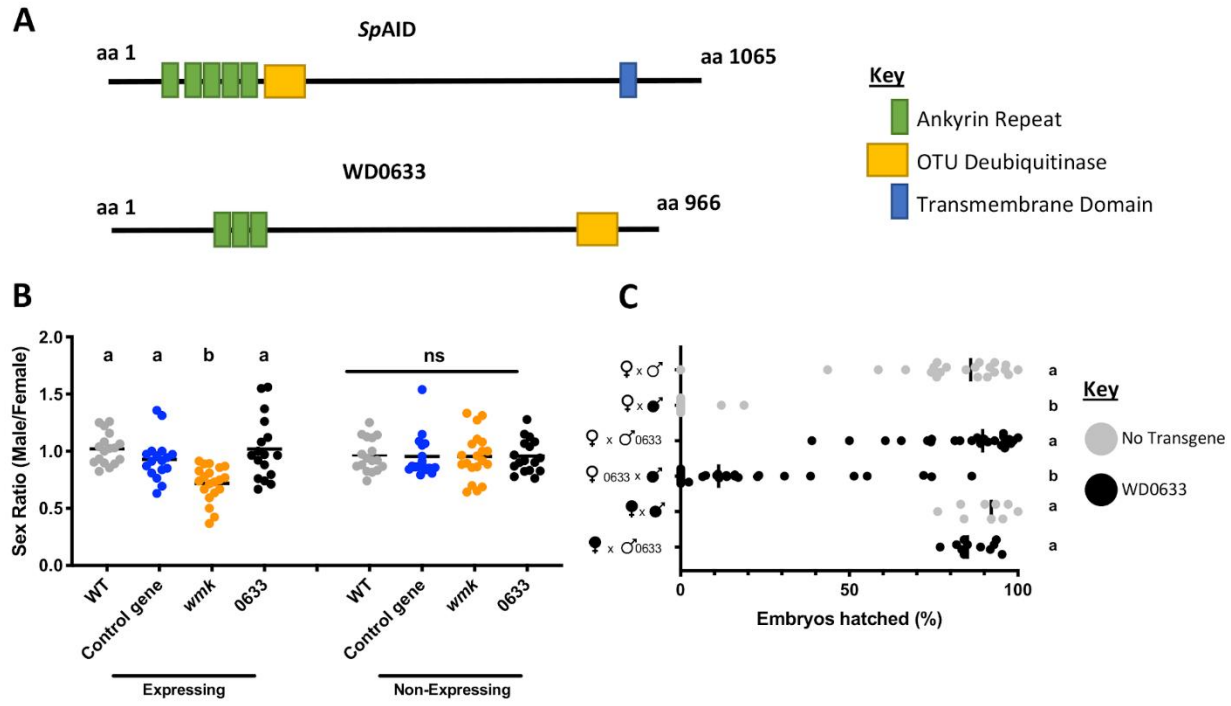


Figure 5-2. Transgenic expression of WD0633 does not recapitulate male killing or CI.

(a) Diagrams of protein architecture using domains indicated from SMART and HHpred databases. (b) Sex ratios of adult flies either expressing (*Act5c*-Gal4) or not expressing (CyO) the indicated genes. Each sample point represents the adult offspring 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. (c) Hatch rate of embryos with infected (filled sex symbol) or uninfected (unfilled sex symbol) flies expressing or not expressing an indicated gene with the *nanos*-Gal4:VP16 gonad-specific driver. Bars represent the median hatch rate. Each dot represents the hatch rate of offspring of a single male and female. Black dots indicate a cross with WD0633 and grey dots indicate crosses without transgenes. Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction.

Interestingly, despite no evidence of a shared ancestry and significantly different protein sequences (1.7% pairwise amino acid identity), the two proteins have similar protein structure predictions (Fig. C-1b). Phyre2 protein modeling[466] indicates that both have putative similarity to the BurrH DNA-binding protein from *Burkholderia rhizoxinica*, a symbiont of *Rhizopus* microspores[467], which is intriguing as Wmk has two predicted helix-turn-helix (HTH) DNA-binding domains[160]. In both cases, BurrH was predicted to be the best model template and results in nearly identical structural predictions for the modeled regions of the two proteins (Fig. C-1b). This suggests they may share similar functions due to architectural similarity despite disparate sequences. This is further intriguing due to the facts that *Sp*AID functions as a male-killing toxin, but WD0633 is not present in all male-killing strains (absent from *w*Rec of *Drosophila recens*), and thus would not be predicted to have male-killing function in *Wolbachia*.

Beyond male killing, WD0633 was also a candidate for CI in an earlier comparative genomic analysis, but it failed to recapitulate the phenotype upon transgenic testing in *D. melanogaster*[464]. However, this was done with a transgene driver that expresses ubiquitously in all tissues rather than specifically in the gonads. In addition, an OTU deubiquitinase domain was previously reported in the CI-causing *cifB* gene[159] based on *in vitro* and yeast studies, though similar effects were not confirmed *in vivo* in flies. Further, WD0633 has multiple motifs and domains that are enriched in type IV secretion system (T4SS) effectors: eukaryotic-like domains (and it is in the EAM), three EPIYA domains, a coiled-coil, C-terminal basicity, and global hydrophobicity (Fig. C-2). Based on these features, the Searching Algorithm for Type IV Effector Proteins 2.0 (S4TE) calculates a high probability of type IV secretion (Table C-1)[468,469]. All of these factors make WD0633 likely to function in the eukaryotic host and thus a particularly

interesting candidate for study in reproductive parasitism. Due to the above reasons, we tested the transgene for recapitulation of male killing with a ubiquitous transgene driver and CI with a gonad-specific driver.

To test for male-killing function, we transgenically expressed WD0633 in *D. melanogaster* flies using a ubiquitous driver (*Act5c*-Gal4/CyO) under the same conditions as previously used to evaluate *wmk*[160]. Transgenic WD0633 expression did not cause a biased sex ratio in adult offspring, similar to a control transgene and WT flies, indicating no recapitulation of male killing in this system. This contrasts with transgenic *wmk* expression, which displayed the expected biased sex ratio when expressed while non-expressing siblings did not have a biased sex ratio (Fig. 5-2b). To test for putative CI function, we expressed WD0633 using the gonad-specific driver, *nanos*-Gal4:VP16, in ovaries and testes of adults. We then crossed adults of the indicated genotypes together (infected, uninfected, or uninfected expressing WD0633 in gonads) and counted the proportion of embryos hatching into larvae as a measure of CI. However, expression in male gonads did not recapitulate CI, and expression in female gonads did not recapitulate rescue (Fig. 5-2c).

Divergent wmk *homologs in* w*Mel do not transgenically recapitulate male killing*

The *w*Mel *Wolbachia* genome *of D. melanogaster* contains the eukaryotic association module (EAM) region in prophage WO, which contains genes with putative eukaryotic functions or homology[20]. The previously-identified *wmk* gene resides in this region, as do several of its homologs. Indeed, some *Wolbachia* genomes contain multiple copies of *wmk* homologs that have apparently arisen by duplication and divergence, or integration of multiple phages. For example, *w*Mel contains four additional homologs, *w*Inn of *D. innubila* contains three, *w*Bor *of D. borealis* contains three, and *w*Bol1b of *Hypolimnas bolina* butterflies contains seven[160]. *w*Mel contains four of these homologs, which share 65-81% pairwise nucleotide sequence identity with *wmk* and result in similar proteins that are all predicted to contain the two helix-turn-helix (HTH) DNA-binding domains annotated in *wmk* (Fig. 5-3)[160]. Although *wmk* is the only *w*Mel homolog shared across all male-killer genomes, we assessed the others in the *w*Mel genome for putative male-killing function in this host as these copies may share the ability to kill males. One of the homologs, WD0508, was previously tested and did not recapitulate the phenotype[160]. Here, we tested an additional three homologs: WD0622, WD0623, and WD0255 (Fig. 5-3b). All are prophage WO

genes, and the first two are in the same prophage WOMelB EAM region as the *wmk* gene, while WD0508 (octomom region) and WD0255 (WOMelA) are in other regions (Fig. 5-1). However, upon transgenic expression using the *Act5c*-Gal4/CyO driver described above, the newly-tested transgenes did not recapitulate a sex-ratio bias, indicating an inability to cause male killing in this system (Fig. 5-3a). Based on this functional analysis, we performed an alignment of the amino acid sequences of Wmk and its homologs and identified 28 amino acid residues unique to Wmk (Fig. 5-3c) that may account for its specific ability to transgenically kill males. These amino acids are spread throughout the protein, and do not yet identify a specific protein region critical for male death (Fig. 5-3d).

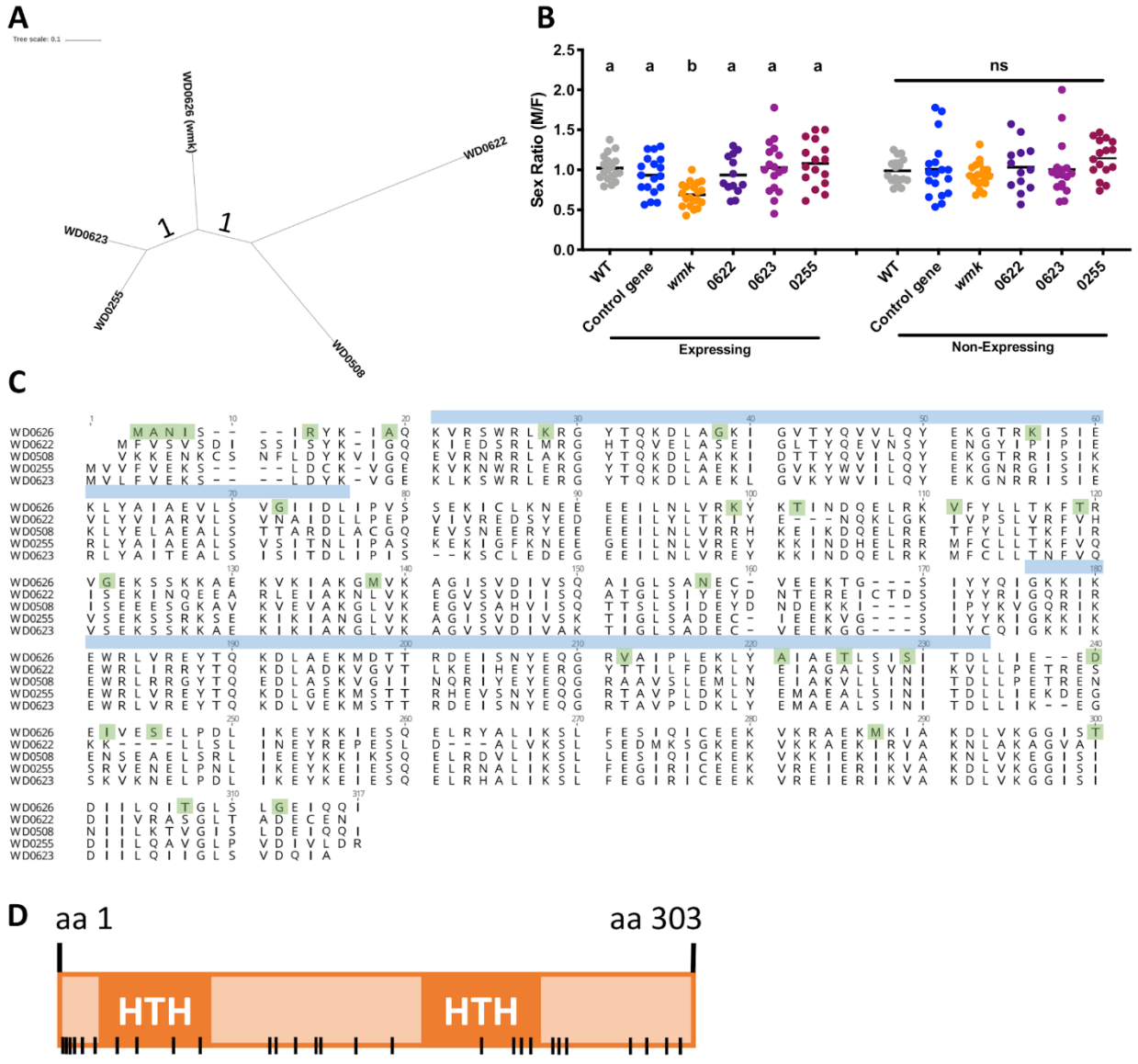


Figure 5-3. Transgenic expression of *w*Mel *wmk* homologs does not recapitulate male killing.
(a) Nucleotide phylogeny of *w*Mel *wmk* homologs. (b) Sex ratios of adult flies either expressing (*Act5c*-Gal4) or not expressing (CyO) the indicated genes. Each sample point represents the adult offspring 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. (c) Amino acid alignment of Wmk (WD0626) and its homologs in *w*Mel. Green highlights indicate amino acids unique to Wmk. Blue boxes indicate the NCBI-predicted position of the HTH DNA-binding domains (above indicated amino acids). (d) Schematic of amino acids unique to WD0626 (Wmk) across the protein sequence, as indicated by black lines. Locations are approximate.

Additional male-killing gene candidates identified via comparative genomics do not recapitulate male killing

In our previous study, we performed a comparative genomic analysis to identify genes associated with male-killer genomes[160]. Among these were *wmk* and *cifA*, the latter of which functions in the induction and rescue of CI. An additional five candidates with a variety of putative functions were not previously tested: WD1243 (putative endonuclease or phospholipase D domain, NCBI conserved domain E=2.22 x 10^{-73}), WD0296 (recombination-promoting PDDEXK family nuclease NCBI conserved domain E=4.50 x 10^{-50}), WD0550 (ankyrin repeat NCBI conserved domain E=1.98 x 10^{-34}), WD0628 (hypothetical protein), and WD0627 (recombination-promoting PDDEXK family nuclease NCBI conserved domain E=2.95 x 10^{-55}) (Fig. 5-4a). Of these genes, one (WD0296) is in a different prophage WO region than *wmk* (WOMelA), and two more (WD1243 and WD0550) are in the *Wolbachia* chromosome of *w*Mel (Fig. 5-1). Of the two that are not in prophage WO regions in *w*Mel, both are phage genes in other strains, including the *w*Bif male-killing strain of *Drosophila bifasciata*[160]. WD0550 contains ankyrin repeats which are abundant in phage WO genes and have been implicated in reproductive parasitism[144,464,470,471]. The other, WD1243, is a putative endonuclease or phospholipase D that has homology to proteins in *Rickettsia* and *Coxiella*, among others, according to a Blastp search. These organisms are common parasites that contain plasmids and other mobile elements[379,472], but ancestral *Ehrlichia* and *Anaplasma* do not contain the gene. In addition, two candidate genes are present in the same EAM region between *wmk* and the *cifA* and *cifB* genes, and therefore might have connections to parasitism due to proximity (WD0627, WD0628). However, when expressed transgenically with

a ubiquitous driver, none of the five additional genes induced a biased sex ratio, indicating that they do not recapitulate male killing when expressed on their own (Fig. 5-4b).

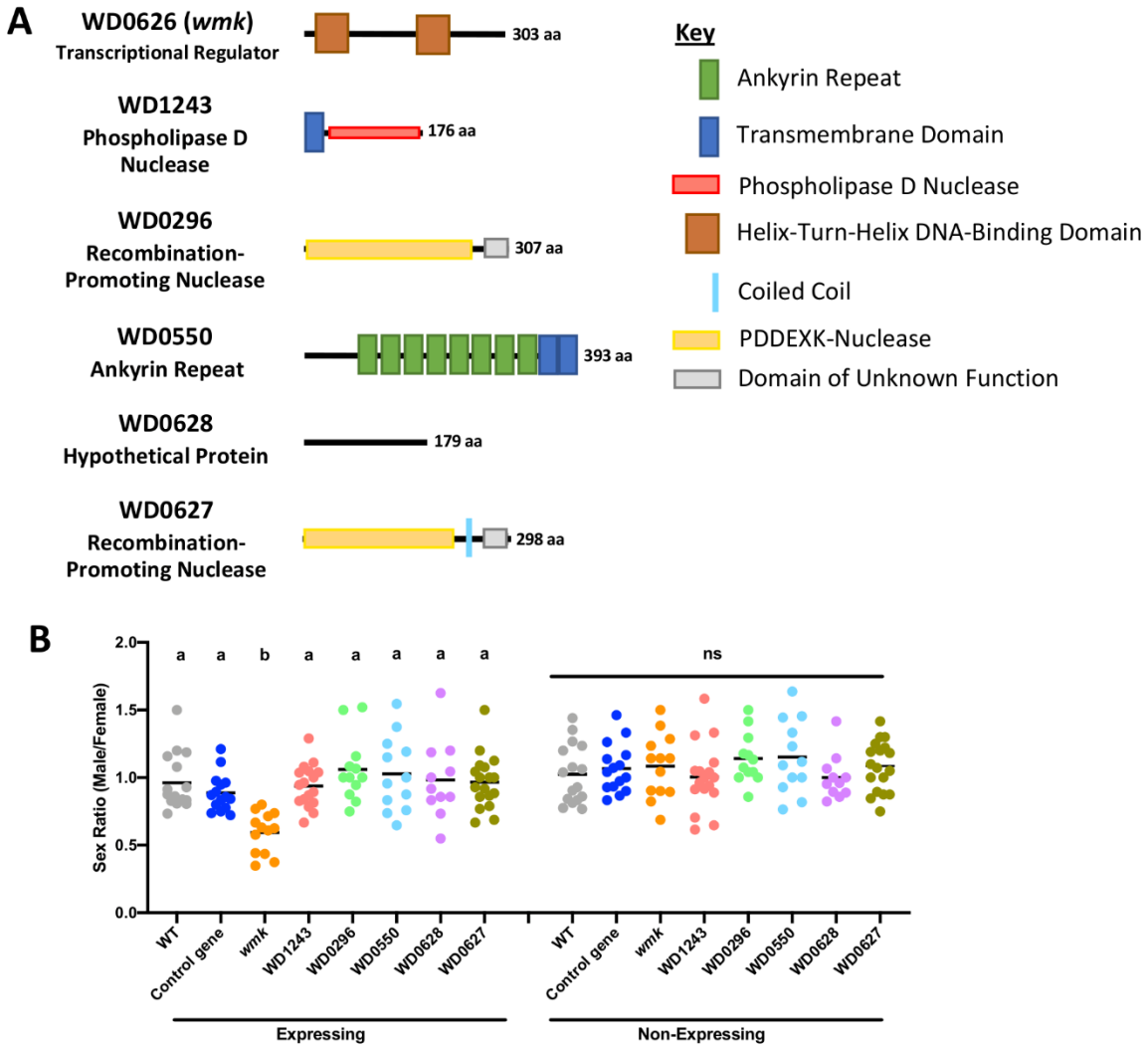


Figure 5-4. Additional male-killing gene candidates does not induce a biased sex ratio with transgenic expression.
(a) Diagrams of protein architecture using domains indicated from SMART[473] and HHpred[474] databases. (b) Sex ratios of adult flies either expressing (*Act5c*-Gal4) or not expressing (CyO) the indicated genes. Each sample point represents the adult offspring 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.

wmk *and* cifA *do not transgenically function together in reproductive parasitism*

Our previous comparative genomic analysis identified *cifA* as a top candidate for male killing[160]. In addition, a similar analysis for CI candidate genes also identified *wmk* as part of a "core" set of CI sequences shared across multiple CI-causing strains[158]. Therefore, both genes appeared as candidates in both reproductive parasitism analyses. Importantly, in the highly reduced prophage region in the *w*Rec male-killing and CI-inducing strain of *Drosophila recens*, there are only 10 prophage WO genes remaining in the region that map to the WOMelB prophage of *w*Mel[402]. In comparison, *w*Mel contains 88 genes in this region. Of the 10 genes remaining in *w*Rec, *wmk* and *cifA* are both included, however there are no additional *wmk* homologs. Indeed, *cifA* and *wmk* commonly co-occur in *Wolbachia* phage WO regions and are located near each other in some male-killer genomes[160]. These co-occurrences in genomes indicate that the two may have similar origins or functions.

Due to their co-appearance in two analyses for CI and male killing, presence in a reduced genome that causes both phenotypes, and close proximity in several *Wolbachia* genomes, we transgenically assessed if they function together by expressing single and dual *cifA* and *wmk*. To test for induction of male killing, we expressed the genes singly or together with the ubiquitous *Act5c*-Gal4/CyO driver and measured sex ratios of surviving adults (Fig. 5-5a). *cifA* does not induce a biased sex ratio on its own, and it neither enhances nor inhibits the ability of *wmk* to cause a biased sex ratio. Therefore, *cifA* is unlikely to play a role in male killing. To test for induction of CI, we performed a hatch rate with the gonad-specific driver *nanos*-Gal4:VP16 to drive expression of each gene singly or together and measured the number of eggs that hatched into larvae to quantify CI induction (Fig. 5-5b). Despite CI induction with either *w*Mel infection or co-expression of the CI-inducing *cifA;cifB* gene combination in males, neither the individual nor co-expressed *cifA* and *wmk* induced CI. This indicates that *wmk* is also likely not involved in induction of CI, as it does not reduce the hatch rate to larvae even when expressed in adult gonads alone or with *cifA*. Only the two CI-inducing genes, *cifA* and *cifB*, are able to fully recapitulate CI induction when expressed together. Therefore, despite the many connections between *wmk* and *cifA*, the data do not support the hypothesis that they function together in reproductive parasitism.

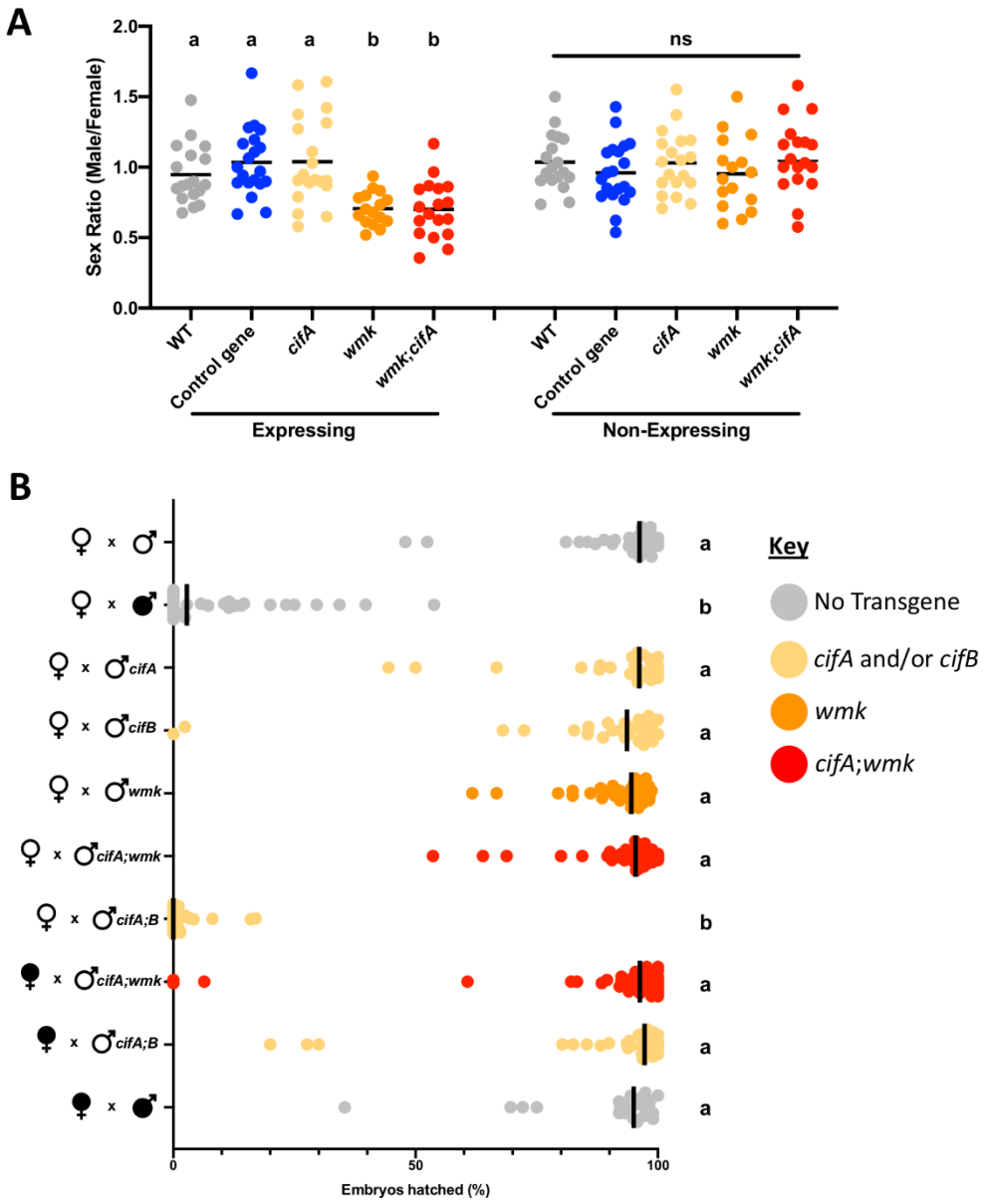


Figure 5-5. Co-expression of *cifA* and *wmk* neither enhances the *wmk* sex ratio bias nor recapitulates CI induction.

(a) Sex ratios of adult flies either expressing (*Act5c*-Gal4) or not expressing (CyO) the indicated genes. Each sample point represents the adult offspring 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. (b) Hatch rate of embryos with infected (filled sex symbol) or uninfected (unfilled sex symbol) flies expressing or not expressing an indicated gene with the *nanos*-Gal4:VP16 gonad-specific driver. Bars represent the median hatch rate. Each dot represents the hatch rate of offspring of a single male and female. Colors indicate the presence or absence of the transgenes as indicated in the key. Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction.

Transgenic expression with alternative transcripts of wmk results in loss of sex ratio phenotype

Previous testing of the *wmk* transgene was performed using its annotated methionine start codon in the NCBI database, which results in a protein of 303 amino acids (aa)[160]. However, additional inspection of the genome identified an alternative start codon (leucine) 9 aa upstream that could putatively produce a 312 aa protein, as bacteria may use non-canonical codons for proteins[475]. To more broadly assess genomes for the presence of alternative start codons, we also analyzed the genomes of four CI-causing and four male-killing *Wolbachia* strains for the presence of any alternative start codons within 100 bp upstream of the annotated start codons (Fig. 5-6a, Table C-2). Indeed, CI-causing strains had between 3 and 5 alternative start codons (mean of 4.5) in this region, while male killers had between 0 and 4 (mean of 1.5). There is additional nuance to this pattern, as *w*Rec is natively a CI-causing strain but can cause male killing in a non-native host, and *w*Bol1-b is the only strain in this group that infects a non-drosophilid host. However, the presence of putative additional start codons is consistent with the hypothesis that there may be expression of alternative transcripts of *wmk* in certain strains or hosts, and it could relate to the presence, or lack thereof, of parasitism phenotypes in a given host. Notably, the WD0508 *wmk* homolog in *w*Mel is annotated with a non-canonical GTG start codon, as are many other *w*Mel genes, so the strain likely expresses at least one homolog of *wmk* with alternative codons. As one potential model, it is possible that expression of *wmk* results in several different transcripts of varying lengths, which is known to occur with some bacterial genes (Fig. 5-6b). Indeed, inspection of publicly-available *w*Mel transcription data shows that young embryos transcribe upstream of the annotated *wmk* start codon in many samples (Fig. C-3a)[387].

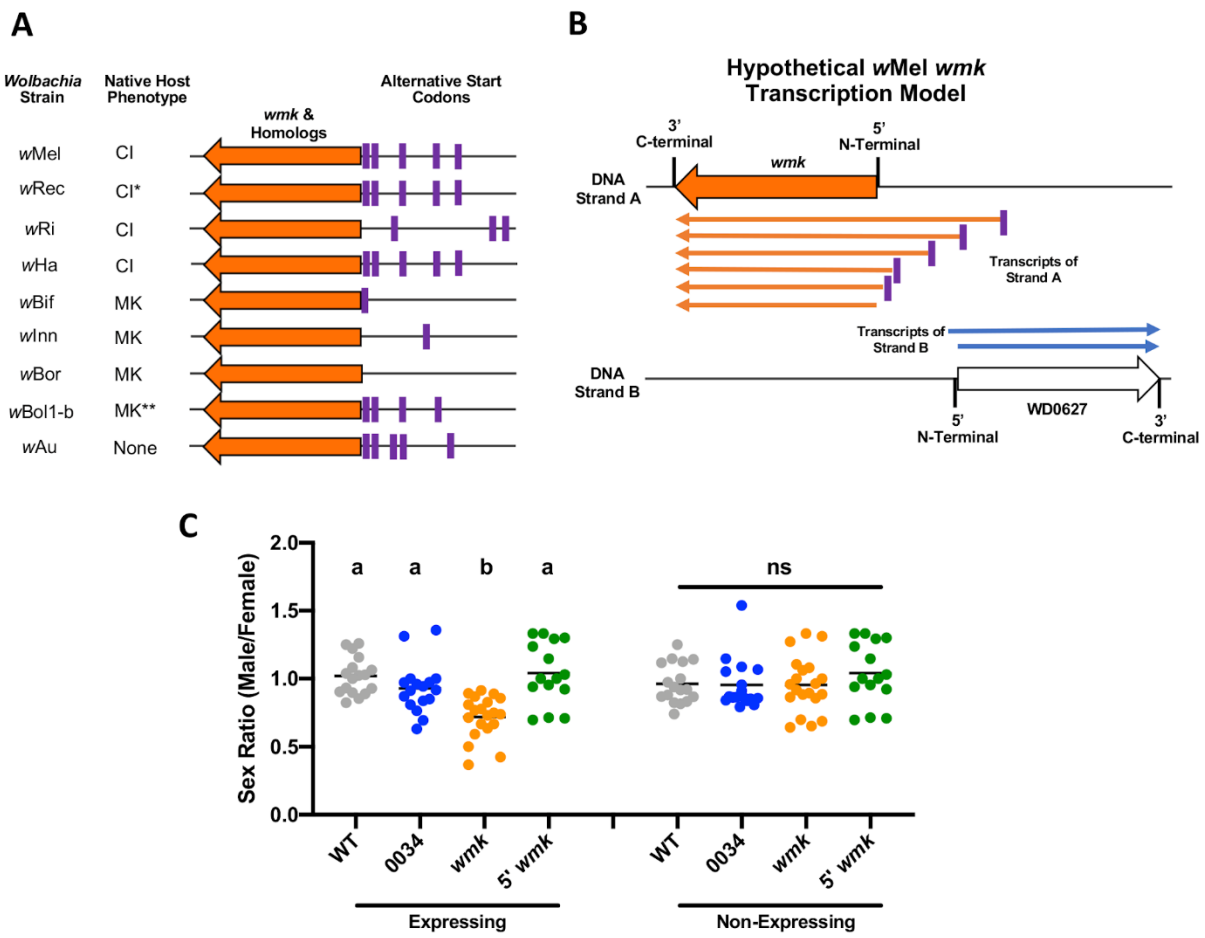


Figure 5-6. Expression of *wmk* with an alternative upstream start codon results in loss of a sex ratio bias.
(a) Diagram of locations of alternative start codons up to 100 bp upstream of *wmk* or its homologs in the indicated strains. Purple stripes indicate the codons, not to scale. **w*Rec causes CI in its native host, but can cause male killing when introgressed into a sister species. ***w*Bol1-b natively infects a non-drosophilid host, the *Hypolimnas bolina* blue moon butterfly, while all other strains in the diagram infect drosophilid hosts. (b) Diagram of hypothetical model where multiple *wmk* transcripts of varying lengths are expressed. (c) Sex ratios of adult flies either expressing (*Act5c*-Gal4) or not expressing (CyO) the indicated genes. Each sample point represents the adult offspring 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.

To functionally test this model, we transgenically expressed an alternative version of Wmk with 9 additional amino acid residues, as this variant had a more commonly-used alternative start codon (TTG) in model prokaryotes compared to others in the upstream region. Although *Wolbachia* start codon usage is unexplored, ATG (81.8%), GTG (13.8%), and TTG (4.3%) are the

most common alternatives in model prokaryotes, and they are used in nearly all cases with other codons occurring less than 1% of the time[475]. Upon codon-optimization and transgenic expression in *D. melanogaster* with a methionine start codon for optimal expression in this eukaryotic organism, the gene loses the ability to cause a biased sex ratio (5' *wmk*, Fig. 5-6c). Though expression of the annotated 303 aa Wmk replicated the biased sex ratio phenotype as in previous experiments, expression with the additional 5' peptide to produce the 312 aa protein with the alternative start codon no longer induced a biased sex ratio. Similarly, expression of a triple epitope HA tag on either the 5' (Fig. C-3b) or the 3' (Fig. C-3c) end also ablates the phenotype. Although the HA tags are small (32 aa) and typically do not interfere with protein function, neither tagged version of Wmk induces a biased sex ratio. These results could indicate that the ends are important for function and that inclusion of even small peptides interferes with protein conformation and thus function, which in turn raises caution for using epitope-tagged proteins in the study of reproductive parasitism. This supports the result that the alternative start codon does not recapitulate the phenotype either. Although it is not clear why longer forms of Wmk lose function, predicted RNA secondary structures of *wmk* and the 5' *wmk* gene tested in Fig. 5-6c show that there are significant predicted structural differences with only the additional 27 nucleotides at the 5' end, and these differences may affect translation (Fig. C-3d).

Discussion

Previous studies identified genes and gene candidates for reproductive parasitism in *Wolbachia* and *Spiroplasma* bacterial endosymbionts[158,160,161,248]; however, continued investigations are necessary to fully resolve the genetics of reproductive parasitism for both basic and applied purposes, especially since the diversity of affected hosts is considerable[6,258]. In particular, the evolution of male-killing by various microorganisms is hypothesized to be the result of convergent evolution of distinct genes due to differences in the timing of lethality (ie, early vs. late male killing) and in sex determination systems across affected hosts (XY and ZW, for example)[258]. Further, *wmk* does not fully recapitulate the phenotype when expressed transgenically. Regarding CI, it also affects a wide variety of hosts, and several diverse phylogenetic Types (e.g., I-IV) of CI genes have been identified[158,463], leaving open the possibility of more. Here we evaluated various gene candidates for CI and male killing that were identified in recent studies on reproductive parasitism genetics (Fig. 5-1). The results and analyses substantiate

cifA, cifB, and *wmk* as the crucial reproductive parasitism genes or gene candidates in phage WO thus far, and the transgenic strains developed here provide a resource to evaluate alternative functions of the tested phage WO genes. Indeed, phage WO genes in the eukaryotic association module are enriched for predicted eukaryotic function or homology, and several of these genes likely interact with the host[20]. There are also many genetic aspects of the endosymbiosis that are not fully explored genetically. These include phage lysis and interactions with the eukaryotic and bacterial host genomes[476], survival of the *Wolbachia* within host cells, other reproductive parasitism phenotypes such as feminization or parthenogenesis, or altered host fecundity[477]. Therefore, despite the lack of evidence for involvement in male killing or CI, there are many other possible phenotypes and functions that these transgenic strains could be used to study.

The WD0633 gene was previously investigated in relation to reproductive parasitism[464,470,478]. The gene is associated with CI genomes[470], contains predicted ankyrin repeats that may mediate interactions with the eukaryotic host[464], is a putative type IV effector (Fig. C-2) predicted to be architecturally similar to the *Sp*AID male-killing toxin (Figs. 5-2a, C-1b), occurs in a mobile element like the other parasitism genes[158,160,248], and is directly adjacent to *cifB* in the EAM[20]. However, it is crucially not present in the assembled genome of *w*Rec[402] that causes both CI in its native host *Drosophila recens* and male killing in the sister species *Drosophila subquinaria*[266]. Despite transgenic assays revealing no role of WD0633 in CI or male killing (Fig. 5-2), it is possible that, like *cifA* and *cifB,* WD0633 cannot function on its own and may require co-expression with another gene, which will be the subject of future investigations. A homolog exists in many *Wolbachia* genomes including the divergent *w*Bif male-killer strain of *Drosophila bifasciata,* but it is missing from the reduced genome of *w*Rec[160,402]. The maintenance of the gene in divergent strains and loss in a strain with an eroded phage may suggest that the gene has an important function in phage WO biology, particularly given the likelihood of secretion and interaction with the host. However, any putative functions or strict associations of WD0633 with active phage WO remain to be assessed. In addition, the similarity between the predicted protein structures of *Sp*AID and WD0633[248] remains a mystery.

Although WD0633 and *Sp*AID have no full homologs outside of *Wolbachia* and *Spiroplasma,* respectively, the OTU domain within *Sp*AID shares weak homology to domains in three *Wolbachia* proteins other than WD0633. Reciprocally, the domain in WD0633 does not have any non-*Wolbachia* homologs. Thus, it remains possible that there was an old gene transfer event

of the OTU domain. *Sp*AID, however, has no known evolutionary connection to sequences within *cifA*, *cifB*, and *wmk*, and there are distinctions between the resulting male killing phenotypes of *Spiroplasma* and *Wolbachia* at the molecular level. For example, *Spiroplasma* results in neural development defects in *D. melanogaster*, while *Wolbachia* shows no such defects in *D. bifasciata*[357]. Given these differences and apparently unique parasitism genes and candidates, it is unlikely that a putative gene transfer from *Spiroplasma* to *Wolbachia* resulted in conferral of a parasitism phenotype.

We also tested the function of *wmk* sequence homologs residing in the *w*Mel genome. Indeed, *w*Mel contains four additional homologs (Fig. 5-3a), along with two partial ones (untested) that are significantly shortened by transposons and contain only one HTH domain each[160]. Not all reproductive parasitic strains contain additional homologs (*w*Rec has only one), and some have fewer copies than *w*Mel (*w*Inn and *w*Bor have three, *w*Bif has one full and one partial). In addition, *wmk* is the only copy in *w*Mel that has direct homologs in all sequenced male-killing strains based on gene synteny and sequence, making it the best candidate of all copies. Under the simplest model, a single *Wolbachia* male-killing gene would be required to kill males. In more complex models, some or all additional copies may together or individually result in a phenotype. As they are similar homologs and since there are multiple copies in many genomes, we tested the *w*Mel homologs for function. WD0508 (71% sequence homology to *wmk*) was previously demonstrated to not kill males transgenically[160]. Here, we show that the full-length homologs WD0255 (81% sequence homology to *wmk*), WD0622 (65% sequence homology to *wmk*), and WD0623 (81% sequence homology to *wmk*) also do not result in a transgenic phenotype (Fig. 5-3b). Their functions, if any, therefore remain undetermined, but there are many possibilities. One is that they do not have a function. Since they vary in number from genome to genome, one or a small number may confer function, and the others may represent copies that have lost or not yet gained a function and may not be maintained over time. Another hypothesis is that they may need to work additively. Phenotypes may emerge or strengthen with expression of multiple copies. For example, the moderate penetrance of transgenic *wmk* could be enhanced by one or more of these homologs. Notably, the CI phenotype can only be transgenically induced by dual expression of the *cifA* and *cifB* genes, and neither can induce the phenotype alone[158]. However, previous dual expression of *wmk* and the adjacent WD0625 gene did not change the phenotype[160]. By singly expressing most candidates in this work (both *wmk* homologs and other candidates), it is possible that their

involvement in a phenotype was missed. However, this is unlikely with the *wmk* homolog candidates since *wmk* is the only full homolog in both the *w*Bif and *w*Rec male-killer genomes that induce high levels of male killing[266]. Alternatively, additional copies (and perhaps *wmk* as well) may have an unidentified function. Indeed, previous work showed the predicted protein structure of Wmk is similar to homologs in other strains despite great amino acid sequence divergence[160]. In addition, *wmk* homologs are ubiquitous in *Wolbachia* genomes despite male killing being a rare phenotype. The maintenance of the gene in non-parasitic strains and conserved predicted protein structure suggest that the Wmk protein may have a pleiotropic function beyond male killing. Indeed, Wmk may have another primary function but can serve as a genetic reservoir for development of male killing in some circumstances. These and additional hypotheses remain to be evaluated in future work.

In addition, it remains unclear why *wmk* has a greater number of divergent copies in some genomes compared to others, but this may be important to its ability to adapt to new hosts or to relaxed selection when host resistance suppresses male killing. If so, multiple copies may be correlated with the proclivity of the host for resistance or suppression in cases where male killing may positively affect *Wolbachia*'s fitness. Indeed, *w*Rec and *w*Bif each have only one full-length homolog, *w*Inn and *w*Bor each have three (there is no documented host resistance in *w*Bif, *w*Inn, and *w*Bor), while *w*Mel has five (and may be resistant). Importantly, *w*Bol1b, the male-killing strain of *Hypolimnas bolina* butterflies contains seven different homologs. The butterfly host is known to develop resistance in many populations, and this resistance fluctuates in a population over time[328,345]. Thus, the presence of many homologs in this strain also correlates with a known tendency for the evolution of host resistance. In other cases where male killing exists with few homologs (such as *w*Inn), it may be that Wmk targets a protein or peptide that could be lethal to the host if mutated, so resistance may be futile as has been previously hypothesized[342]. Further, these three copies are in the same phage WO region with a synteny that matches many other strains, and it is therefore likely that these three copies existed before insertion into these strains[160]. In contrast, *w*Mel and *w*Bol1b have other *wmk* copies not shared by most strains, so they may have uniquely arisen in these prophages. Further, some additional copies in the genomes are the simple result of multiple phage WO insertions with their own version of *wmk* in one *Wolbachia* genome. Few copies may therefore correlate with less host resistance while more may correlate with greater resistance, which is the basis of a copy number-host resistance hypothesis. Notably, we can use

the negative transgenic expression results from these additional copies to narrow down important residues for the male death phenotype induced by Wmk (the only functional copy thus far). There are 28 residues unique to Wmk, which alone or in combination may be important to the onset of the phenotype (Fig. 5-3c). While they are not clustered in any region of the protein, they may help narrow down points of interactions with putative host targets to be assessed in the future.

Next, we functionally evaluated male-killing candidates identified in our previous comparative genomic analysis (Fig. 5-4)[160]. These candidates were identified by looking for genes shared across the *w*Bif, *w*Inn, *w*Bor, and *w*Rec male-killer genomes, among other criteria. Seven were identified, among which were *wmk* (evaluated previously) and *cifA* (evaluated previously and in Fig. 5-4). Five others remained (Fig. 5-4a). Upon transgenic expression, none induced a biased sex ratio (Fig. 5-4b). They therefore do not recapitulate the phenotype in this system. This does not rule out their role in natural parasitism phenotypes conclusively, but it is unlikely they are parasitism genes. It is possible that the transgenic system is not able to fully induce male killing by these genes or that they must be expressed in conjunction with another gene. However, there is no additional evidence (such as homology to toxin-antitoxin systems, etc.) to suggest they function together with *wmk* or any other gene. Of the five genes, WD0627 and WD0628 may be of the most interest for further parasitism research. Their adjacent position does not necessarily suggest they function with *wmk* since they are on the opposite DNA strand, which is atypical of co-transcribed genes, although the possibility remains. They are, however, located between *cifA* and *wmk* in the EAM region and are of interest due to their general proximity with these genes.

Given the co-associations of *wmk* and *cifA* in genomics analyses and in physical relation to each other in some divergent genomes (such as *w*Mel and *w*Bif)[158,160], we hypothesized that the two genes may function together. However, transgenic assays did not demonstrate additive or epistatic effects when they are co-expressed (Fig. 5-5). Therefore, their relationship with each other, if any, remains unsolved. There are a few possibilities, which are not all mutually exclusive. One is that their co-localization is coincidental due to putative shared origin. Similarly, they may have close proximity (within ~5 kb of each other) by chance and not common origin. Another possibility is that *Wolbachia*'s fitness in a host is contingent upon expressing multiple parasitism phenotypes. If environments are rapidly fluctuating, and these genes are subject to genotype by environment interactions, multiple genes may enable *Wolbachia* to spread in such circumstances without the cost of relaxed selection on the unexpressed genes. For example, male killing may not

be penetrant in high temperatures, but CI might still manifest, allowing the bacteria to proliferate in these conditions[262]. Moreover, if one phenotype is ablated due to genetic mutation or host resistance, then the other provides a backup. Two or more genes expressing different phenotypes could function as bet-hedging to benefit *Wolbachia* in complex ecological and environmental scenarios. These or other premises remain to be tested.

As genes beyond *wmk* were evaluated and did not recapitulate CI or male killing, the question remains why transgenic *wmk* cannot fully induce a male-killing phenotype and why *w*Mel does not naturally kill males in *D. melanogaster*. It is still possible that other genes may be involved that have not yet been discovered, however, we have now evaluated all top candidates identified thus far. In addition, we previously used many different transgenic drivers to test the premise that different expression levels are required for the phenotype[160]. However, increasing expression by an order of magnitude does not result in a change, indicating expression levels likely do not underlie the partial phenotype[160]. The lack of a full transgene phenotype remains unsolved, as does the lack of male killing via natural infection. One possibility we explored was that alternative transcripts of *wmk* might underlie a natural lack of phenotype. Inspection of the upstream DNA identified several alternative start codons in addition to the annotated methionine codon that are possibly used in embryos (Figs. 5-6a, 5-6b, C-3a). Transgenic expression of the most likely alternate transcript yielded no phenotype (Fig. 5-6c), supporting the hypothesis that alternative transcripts could underlie phenotypic differences.

There is an imperfect correlation that the male-killer genomes do not have as many alternative start codons as the CI genomes (Fig. 5-6a). The notable exceptions are the multipotent *w*Rec as well as *w*Bol1b, which infects a butterfly host of ZW sex determination and may not be affected by the same male-killer toxin as drosophilids (XY sex determination)[253]. Notably, inspection of the three most commonly used prokaryotic alternative start codons in each of the genomes depicted in Fig. 5-6a reveals that the *w*Bif and *w*Bor male-killer genomes contain none of these three common codons upstream of the annotated start. Only less-common codons are present. The *w*Inn homolog does contain a GTG codon upstream, however, in model prokaryotes this is much more commonly used than the annotated ATA codon in this strain[475]. Thus, the GTG may be the main start codon, and the ATA codon could represent a mis-annotation by the Glimmer gene prediction program. All others (the CI strains, the non-parasitic strain, and *w*Bol1b) contain at least one of the top three most used codons upstream of the annotated start. Therefore, of the

three, natural male-killers in drosophilids, it is possible that none express alternative transcripts or a very low amount. The CI-causing and non-parasitic strains, however, may express higher numbers of longer transcripts that impede function. Indeed, addition of other elements such as an HA tag on either end of the protein (that do not traditionally interfere with protein activity), ablates the phenotype as well (Figs. C-3b and C-3c). This suggests that alternate *wmk* transcripts, which may appear commonly in genomes with certain start codons, may result in loss of function. This therefore represents a hypothesis underlying the natural inability of *w*Mel to kill males in its host, as it may express non-functional forms of the protein. Transgenic expression, however, would not encounter this issue as the transcript has only one start codon optimized for host expression. The potential difference in the transcript lengths is one possible reason transgenic expression of *wmk* results in a phenotype while natural expression does not.

The basis of the difference in phenotype from the transcripts is unclear, but may lie in potential differences in transcriptional or translational speed affecting the amount of protein. The two phenotypes from the two transcripts could potentially result from different RNA secondary structure impeding protein translation or altering protein folding (as in Fig. C-3d). Indeed, experiments in *E. coli* demonstrate that the first five to ten codons in an mRNA transcript greatly determine mRNA folding at the translation start, and this region of mRNA structure is the primary determinant of translation rate[479]. Different resulting translation rates are the proposed basis of selection for non-canonical start codons, as codons would be selected based on their effect on translation. Notably, several of the upstream codons are conserved across several strains, in terms of both codon sequence and location. This conservation in a non-coding region supports the sequences having putative functional importance, potentially as alternative start codons as tested here or as alternative promoters to the gene. Future work should compare and contrast the transcripts and putative promoters of various strains to further assess these hypotheses *in vivo*. It remains unclear why *w*Rec would induce male killing in a sister species with the presence of a common alternate codon; however, it is notably unable to kill males in all *D. subquinaria* strains[266]. Therefore, we do not yet understand why the transgenic phenotype is weak, but we present a new hypothesis for the difference between the transgenic and (lack of) native phenotype in which the transgene only expresses the transcript that leads to a sex ratio bias, while the native strain expresses some number of non-functional transcripts.

Here, we evaluate and present evidence on the role of many *w*Mel genes in CI, male killing, or both. The hypotheses that these other genes are involved in their tested phenotypes are not supported by the data. Notably, most of the genes were tested singly, and it is possible that they work together to induce a phenotype. Further, *Wolbachia* strains can induce weak vs strong phenotypes[480,481] or CI vs male-killing[266,341], depending on different factors including host background. Thus, testing candidate genes in other host genetic backgrounds will be an important future direction that may yield new or different results. Previously, we tested dual *wmk*;WD0625 expression as they had the potential to be co-expressed, but this did not result in a change in phenotype[160]. The only two gene candidates within this work that were anticipated to be linked were *wmk* and *cifA* due to the reasons explained above, however future work may create new dual-expressing lines of different gene combinations to determine if additional genes function in parasitism phenotypes. The results also generated several new hypotheses and analyses relating to the connection between *Spiroplasma* male killing and *Wolbachia*, the origin of multiple copies of *wmk* in a genome, Wmk residues critical for protein function, the correlation between *wmk* and *cifA*, and a putative transcriptional basis to some of the complexities of *Wolbachia* genotype and phenotype that could be tested with future work. This work not only advances our understanding of the role of phage WO genes in eukaryotic host biology, but will also spur new research into the unique genetics of this symbiosis.

Materials and Methods

Fly strains and transgene constructs

D. melanogaster strains used in this study include several available at the Bloomington *Drosophila* Stock Center: *Act5c*-Gal4/CyO (BDSC 3953, ubiquitously-expressing zygotic driver), y^1w* (BDSC 1495, *Wolbachia*-infected), tetracycline-treated y^1w^* (uninfected), the WT background line of genotype y^1w^{67c23}; P[CaryP]P2 (BDSC 8622), and *nanos*-Gal4:VP16 (BDSC 4937, gonad-specific driver). In addition, transgene constructs described in our previous publication on wmk include WD0034 (control gene) and WD0626 (*wmk*), both of which were codon-optimized for *Drosophila* expression and synthesized by GenScript Biotech (Piscataway, NJ) on a pUC57 plasmid, cloned using standard molecular biology techniques into the pTIGER pUASp-based vector for germline expression that integrates using PhiC31 integrase, and inserted into the BDSC 8622 background line by Best Gene, Inc (Chino Hills, CA), with transformants

selected based on w^+ eye color. In addition, previously described constructs include WD0632 (*cifB*, insert line BDSC 8622) and WD0631 (*cifA*, y^1 w^{67c23}; P(CaryP)attP40, CytoSite 25C6 insert line from BestGene), which were generated with the same process. The dual WD0631;WD0632 (*cifA;cifB*) line was generated using standard introgression of the two lines. Here, we also describe new transgene constructs. The 3'HA WD0034 (control gene) and 3'HA WD0626 (*wmk*) lines were made with the same process as the above constructs, but were cloned onto the pTIGER-3'HA vector, which includes in additional 3' triple HA epitope. WD0622 (BDSC8622), WD0623 (BDSC8622), WD0255 (BDSC8622), WD1243 (BDSC8622), WD0296 (BDSC8622), WD0550 (BDSC8622), WD0633 (BDSC 9736, y^1w^{1118}; pBac(y[+]-attP-9A)VK00018 insert line), WD0627 (BDSC8622), and WD0628 (BDSC 9736), 5' HA WD0034 (control gene, triple HA epitope, BDSC8622), 5' HA WD0626 (*wmk*, triple HA epitope, BDSC8622), and 5' WD0626 (*wmk*, 5' alternative start codon 9 aa upstream, BDSC8622), were all generated via *Drosophila* codon optimization and gene synthesis followed by cloning into the pTIGER plasmid by GenScript Biotech (Piscataway, NJ), and subsequent injection and integration of the plasmid into their respective background lines by Best Gene, Inc (Chino Hills, CA), with transformants selected based on w^+ eye color. The dual WD0626;WD0631 (*wmk;cifA*) line was generated using standard introgression of the two lines.

Fly maintenance

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[160]. Briefly, twenty 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 set up in vials with CMY media. 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). Any vials with fewer than 50 adult offspring were removed from the analysis, as this indicates either poor egg laying or abnormally low egg hatching. The number of adult offspring per vial ranges from 50-170, there is a mean of 120, and a standard deviation of 27.

Hatch rate assays

To assess the effect of transgene expression on embryo hatch rates (measurement of CI), hatch rates were performed as previously described[160]. Briefly, adult, virgin paternal and maternal grandmother females were aged 9-11 days before crossing with non-virgin, non-age-controlled grandfather males of the desired genotype. All uninfected mothers and fathers were derived from crosses between grandmother nanos-Gal4:VP16 crossed to either tetracycline-treated y^1w^* or a transgene grandfather. All infected mothers and fathers were derived from crosses between y^1w^* (infected) grandmothers crossed to tetracycline-treated y^1w^* grandfathers. All steps on the maternal side were started 7 days prior to the equivalent step on the paternal side. Mothers were aged 5-7 d and fathers were aged 0-24 h before crossings. Fathers in hatch rates are younger than mothers due to the established CI aging effect, where CI gets weaker as a male ages[386]. The mothers and fathers were crossed in single pairs in 8 oz. round-bottom Drosophila bottles covered with a grape-juice agar plate (created as previously described) with a small smear of yeast paste and tape to hold it down, with 32-48 individual crosses per genotype. The bottles were stored with the agar plate down at 25°C overnight (~16 h) and grape juice agar plates were swapped for fresh plates with yeast. The flies were then again stored with the agar plate down at 25°C for 24 h. Plates were then removed, and parents were discarded. The plates were kept at 25°C except during counting. The embryos on the plate were counted immediately upon removal and the number hatched were counted again at 36 h. The hatch rate was calculated as the percent hatched out of the total number laid. Any plates with fewer than 25 embryos per mating pair were removed from the analysis, as this indicates poor egg laying. The number of embryos per plate ranges from 25-125, there is a mean of 50, and a standard deviation of 20.

Domain and motif analyses

Protein domains were identified first by using the protein sequences from the NCBI database and running them through SMART (Simple Modular Architecture Research Tool, http://smart.embl-heidelberg.de/)[473] to identify and annotate protein domains. The images produced by this software were used as the basis for Figs. 5-2a and 5-4a. Additional domains were added if identified in subsequent analysis, described here. In addition, the amino acid sequences were run through HHpred (https://toolkit.tuebingen.mpg.de/#/tools/hhpred)[474] to confirm SMART-identified domains and identify additional domain structures in each protein. It was run with default parameters and the following databases: SCOPe70 (v.2.07), COG/KOG (v1.0), Pfam-A (v.32.0), and SMART (v.6.0). Domains were included if they were predicted by SMART and/or contained probabilities greater than 90% in HHpred. The only exception was *Sp*AID, as the ankyrin repeats dominated the results. A second HHpred analysis was done on only protein sequence after the repeats, which identified the previously-reported OTU domain with 94.25% probability.

T4SS motif identification

Type IV secrection system (T4SS) effector motifs were identified using the S4TE (Searching Algorithm for Type IV Effector proteins, v.2.0, http://sate.cirad.fr/). S4TE is a suite of online bioinformatics tools that analyzes protein sequences for 14 characteristics associated with effectors (such as homology to effectors, eukaryotic domains, subcellular localization signals, etc.) and scores the proteins[469]. Those above a threshold value are predicted to be secreted. The analysis was performed by selecting the "*Wolbachia* endosymbiont of *Drosophila melanogaster*" NCBI genome option and running S4TE 2.0. Out of 1195 total proteins, 148 were above the threshold score of 72, and WD0633 scored highest with 246 based on its characteristics and is therefore likely to be secreted. Fig. C-2 is the output figure by the program.

Alternative start codon identification

Alternative start codons were identified using Geneious Pro 2019.2.1. The gene sequence of each *wmk* homolog from *w*Mel, *w*Rec, *w*Bif, *w*Inn, and *w*Bor with the additional intergenetic sequence between *wmk* and WD0627 homologs were analyzed. ORFs were identified with the Find ORF function in Geneious with the following parameters: a minimum size of 300 nt;

including interior ORFs and continued outside sequences; bacterial genetic code; and CTG, ATC, TTG, ATA, ATG, ATT, and GTG as alternative start codons (default codons of the program). The identified codons are listed in Table C-1.

Phylogenetic analyses

For Fig. C-1a, a Blastp search was done with either the OTU domain of *Sp*AID identified in NCBI (residues 343-431) or residues 1-342, 432-732, and 733-1065. Only hits with E-values less than 1×10^{-5} were included in the analysis. Hits to the latter three residue regions were all *Spiroplasma* sequences and were not included in further analysis. Results for the OTU domain included three regions from *Wolbachia* sequences and five *Spiroplasma* sequences. These sequences were exported and uploaded to Geneious Pro v.2019.2. The sequences were aligned using MUSCLE[409] and indels were deleted. The alignment was imported to ProtTest v.3.4.2[482,483] and the AICc-corrected prediction for best model was cpRev. The MrBayes[484,485] plugin of Geneious was used to generate a tree using cpRev as the model and the consensus tree was exported and imported to iTOL v4.4.2[486], where the final display tree was generated. The same process was used to generate the phylogeny in Fig. 5-3, except nucleotide sequences were used and JModelTest v.2.1.10[483,487] predicted JC as the AICc-corrected best model, which was used in the construction of the tree using MrBayes.

Protein alignment

The protein alignment in Fig. 5-3c was generated by using a MUSCLE alignment of all sequences in Geneious Pro v.2019.2. Discrepancies in sequences were highlighted and unique WD0626 (Wmk) sequences were marked manually.

Statistical analyses

All statistical analyses for sex ratios and hatch rates were performed using GraphPad Prism 8 software. For sex ratios, a non-parametric Kruskal-Wallis one-way ANOVA followed by Dunn's test of multiple corrections was applied to all gene-expressing categories, followed by the same test but on all non-expressing categories. For hatch rates, a non-parametric Kruskal-Wallis one-way ANOVA followed by Dunn's test of multiple corrections was applied to all crosses.

Acknowledgements

We would like to thank Sarah Bordenstein for helpful discussion on the data and identification of an alternative start codon. 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.

Chapter 6

Assessment of transgenic homologs of a male-killing gene candidate

Contributing Authors: Jessamyn I. Perlmutter, Jane E. Meyers, and Seth R. Bordenstein

Abstract

Wolbachia are the world's most widespread bacterial symbionts, and primarily infect arthropods and nematodes. These obligate intracellular bacteria are maternally transmitted and facilitate their spread by hijacking arthropod host reproductive processes. "Reproductive parasites" such as *Wolbachia* have profound evolutionary impacts on arthropod host populations and are at the forefront of vector control strategies around the globe. Identifying the genes and mechanisms that underlie parasitism phenotypes are critical to better assess the impact of these symbionts and their use in the field. One such parasitism phenotype is male killing, whereby infected males are selectively killed. We previously identified and assessed a gene candidate for *Wolbachia*-induced male killing and found that it recapitulated many aspects of the natural phenotype when expressed transgenically in *Drosophila melanogaster* embryos. Here, we transgenically express homologs of *wmk* from diverse male-killing *Wolbachia* strains in *D. melanogaster*. We find that while distantly-related homologs do not induce a biased sex ratio, homologs similar to *w*Mel *wmk* in sequence either kill some males or kill both males and females. In addition, we find that there is a complex relationship between *wmk* genotype and phenotype, where even synonymous mutations may result in significant changes to RNA secondary structure and resulting phenotype. The current work provides important insight into the potential relationship between *wmk* and male killing as well as the utility of transgenics in functional testing of endosymbiont phenotypes.

Introduction

Wolbachia are maternally-transmitted, obligate intracellular bacteria that infect the germline tissues of many arthropod species around the world[23]. To facilitate their spread through the matriline of a host population, these microbes hijack host reproductive processes through various reproductive parasitism phenotypes. One such phenotype is male killing, in which sons of infected females are selectively killed[258]. This can have profound impacts on host population evolution, including the possibility of extinction or sex role reversal[176], and host nuclear genome changes to resist the phenotype[280,345]. It is also a theoretical method for control of arthropod pests

and disease vectors[373]. To advance our understanding of the impact that male killing has on host populations and to better assess it as a putative tool for arthropod control, it is important to determine the genetic basis and molecular mechanism of male killing

Recently, we identifed a prophage WO gene in *Wolbachia* that is a candidate for male killing[160]. This gene, termed *WO-mediated killing* (*wmk*), contains two helix-turn-helix (HTH) DNA-binding domains and recapitulates many aspects of male killing when transgenically expressed in *Drosophila melanogaster*. The transgene results in a female biased sex ratio due to death of male embryos, and male embryos exhibit DNA defects similar to those in natural infection. Additionally, expression leads to DNA damage that coincides with dosage compensation. We concluded that *wmk* is a top candidate for reproductive parasitism based on functional and comparative genomic analyses[160]. However, the *wmk* transgene was characterized from the *w*Mel *Wolbachia* strain of *D. melanogaster* that is not known to naturally cause male killing. Absence of male killing by the *w*Mel strain in *D. melanogaster* is not conclusive, as phenotype switching is common whereby a strain causes a phenotype in one host, but a different phenotype in another host[266,267,341]. To further assess *wmk* as a putative male-killing gene, we sought to transgenically develop *Drosophila melanogaster* strains that express homologs of *wmk* from strains that are known to kill males. Here, we present results that transgenic expression of *wmk* homologs from non-native strains in *D. melanogaster* results in complex phenotypes that may be intricately tied to transcriptional-level factors.

Results

To determine if *wmk* homologs from other strains induce a biased sex ratio when expressed in fly hosts, we transgenically expressed several codon-optimized homologs from known male-killer strains in *Wolbachia* including the *w*Bol1b strain from *Hypolimnas bolina* butterflies[247], the *w*Bif strain from *Drosophila bifasciata* flies[356], the *w*Caub strain from *Cadra cautella* moths[341], and the *w*Inn strain from *D. innubila* flies[264] (same sequence as *w*Bor strain from *D. borealis* flies[240]). These all included the male-killer *wmk* sequences more distantly related to *w*Mel *wmk*[160]. While transgene expression of *w*Mel *wmk* induced a biased sex ratio as expected (about a third of expressing males die), no other transgenes resulted in a biased sex ratio in expressing hosts, indicating that they cannot recapitulate male killing when expressed in *D. melanogaster* under the driver and insertion sites used here (Fig. 6-1).

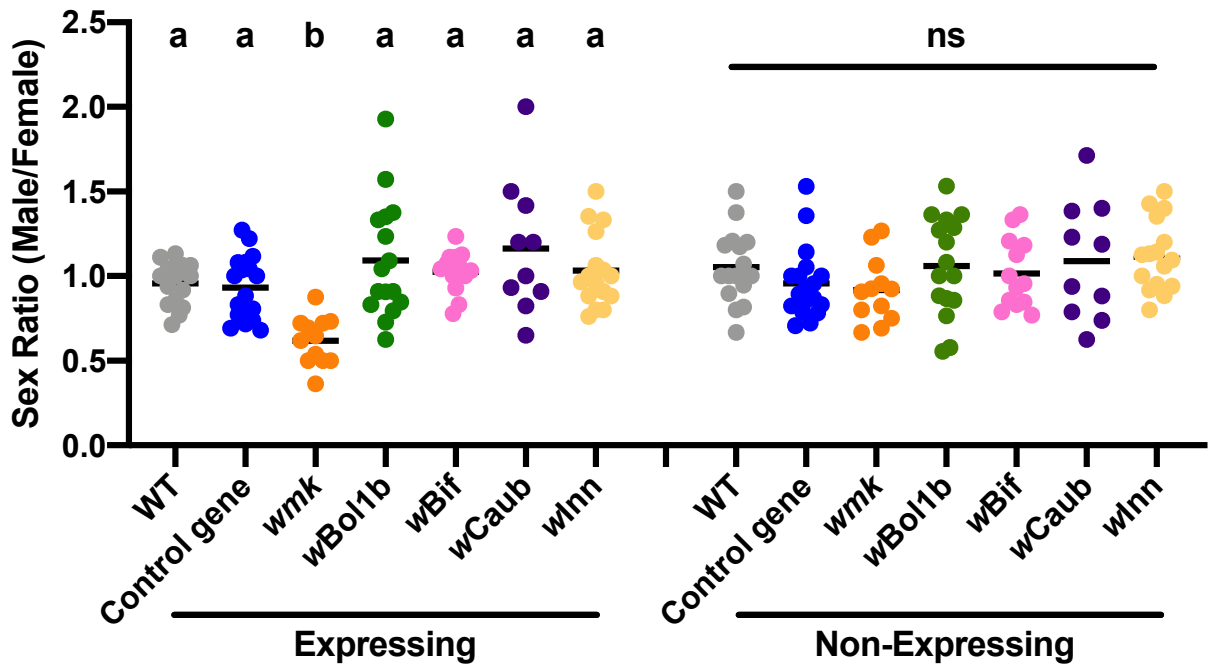


Figure 6-1. *wmk* homologs from male-killing strains distantly-related to *w*Mel do not induce a biased sex ratio in *D. melanogaster*.
The graph shows sex ratios of adult flies either expressing (*Act5c*-Gal4) or not expressing (CyO) the indicated genes. Each sample point represents the adult offspring 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.

We also tested *wmk* homologs from more similar strains to *w*Mel, including *w*Suzi of *D. suzukii* (not a known male-killer, but interesting due to its pest status[488]), and *w*Rec of *D. recens* (kills males in sister species, *D. subquinaria*[266]). The *w*Suzi homolog has three nucleotide differences compared to the *w*Mel homolog, but has the same amino acid sequence. The *w*Rec homolog has three nucleotide differences, one of which is non-synonymous, making the protein a single amino acid different from *w*Mel *wmk* that is in the first HTH DNA-binding domain. All transgenes were codon optimized. We anticipated that both would still be able to induce a biased sex ratio comparable to *wmk* when transgenically expressed in *D. melanogaster*. Unexpectedly, transgenic expression resulted in death of all expressing flies, both male and female. 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* (Fig. 6-2a). To assess potential differences in gene expression, we

measured the gene expression levels of the transgenes in embryos 4-5 h AED (after egg deposition, when wmk kills males), and confirmed that gene expression levels are not significantly different across *w*Suzi, *w*Mel, *w*Rec, or HA tag *w*Mel transgenes, suggesting that RNA levels do not account for phenotype differences (Fig. 6-2b). In the absence of an antibody, we were unable to tell if there are protein abudnance differences. As one alternative hypothesis, we considered that expression of the *w*Suzi and *w*Rec homologs may affect expression of another host gene in a way that is lethal to all hosts. Indeed, in *D. melanogaster* flies artificially expressing *msl-2*, both males and females are susceptible to *Spiroplasma poulsonii*-induced killing[358]. We therefore hypothesized that *msl-2* expression may have been increased, leading to killing of both males and females as well. To measure this, we quantified *msl-2* mRNA levels in embryos expressing *w*Mel *wmk* (sex ratio bias), *w*Suzi *wmk* (all expressing hosts die), or *w*Bif *wmk* (a homolog with no phenotype). However, *msl-2* levels were comparable across all groups, suggesting no significant difference in *msl-2* expression in *w*Suzi *wmk*-expressing embryos (Fig. 6-2c).

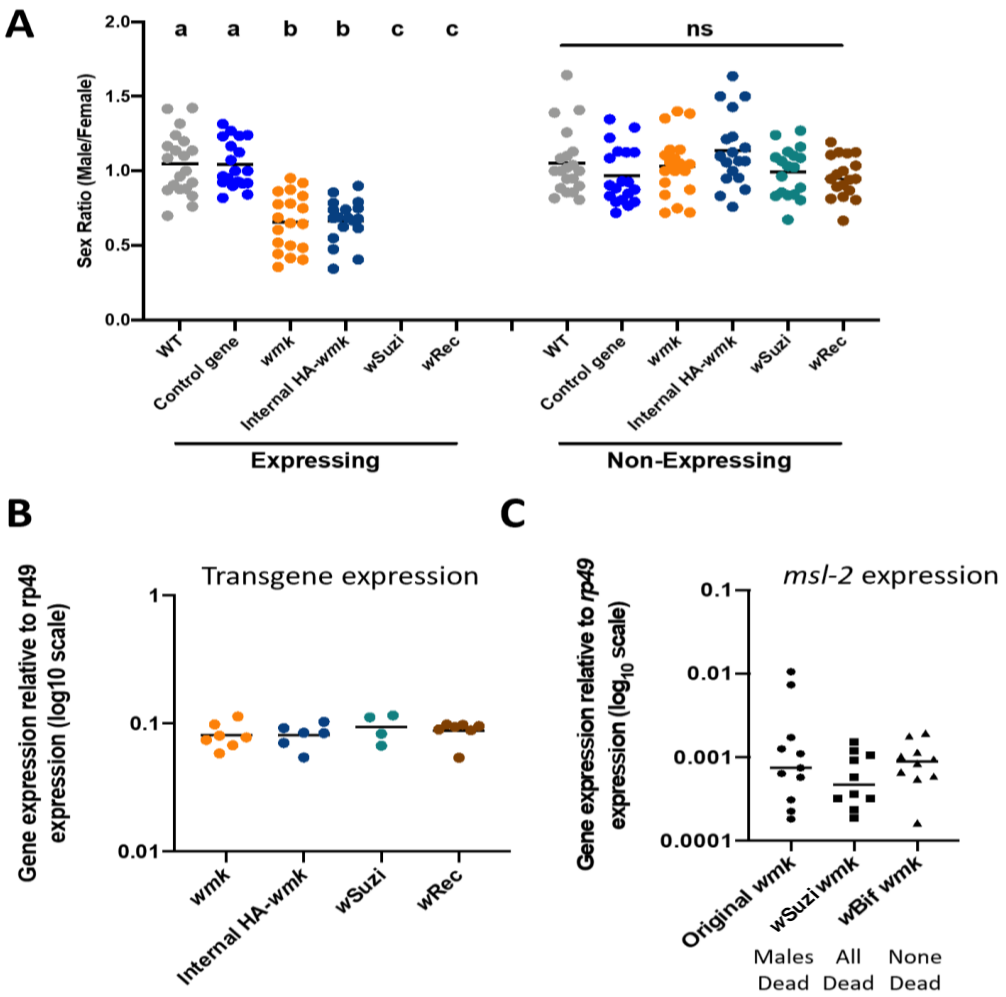


Figure 6-2. *wmk* homologs most similar to *w*Mel *wmk* exhibit a range of transgenic phenotypes in *D. melanogaster*.
(a) Sex ratios of adult flies either expressing (*Act5c*-Gal4) or not expressing (CyO) the indicated genes. Each sample point represents the adult offspring produced by a replicate family 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 (up to 3 survivors). (b) Gene expression in embryos 4-5 h AED of each indicated *wmk* transgene from (a), 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. (c) 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.

Next, we considered that other transcriptional differences may account for phenotypic variation from each genotype. To assess this, we compared the predicted mRNA secondary structures of several of the genes and determined that despite few to no amino acid level changes

from one protein to another, the RNA secondary structures were predicted to be substantially different (Fig. 6-3). Indeed, there are some structural differences comparing native *w*Mel *wmk* to transgene *w*Mel *wmk* (Figs. 6-3a and 6-3b) including the position of the start codon (indicated by blue arrows) and number and size of various loops and hairpins. The *w*Suzi and *w*Rec homologs also varied from the other RNA secondary structures, but the HA-tagged *w*Mel *wmk* had a surprisingly similar structure to the *w*Mel *wmk* RNA despite having a large sequence (111 nucleotides) inserted into the middle of the gene (Figs. 6-3c to 6-3e). Although there are few major structural differences that clearly differentiate transgene *w*Mel *wmk* and the HA tag *w*Mel *wmk* from the others, they both have fewer loops with the start codon at the middle of the structures, while the others do not (Figs. 6-3b and 6-3e). Thus, it is possible that RNA secondary structure contributes to the resulting differences in phenotype.

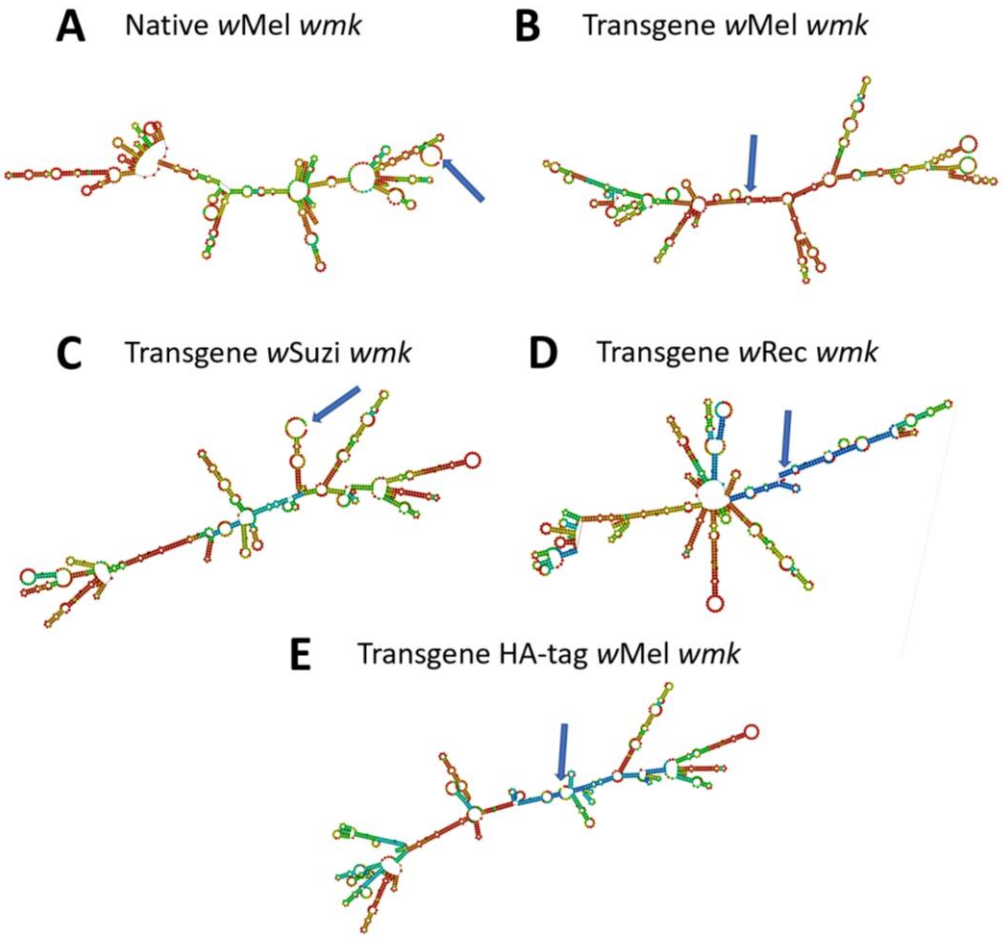


Figure 6-3. Predicted RNA secondary structures of the indicated native *Wolbachia* or transgenes exhibit large differences in structure.

All structures are of the indicated native *Wolbachia* gene or transgene. Blue arrows point to the start codon of each gene.

Indeed, we have previously provided evidence that transcriptional changes in *wmk* expression may contribute to differences in male-killing abilities between *Wolbachia* strains or hosts[369]. In this earlier work, we demonstrated that some strains contain alternative start codons upstream of the annotated start for *wmk* and that embryos do express transcripts in this upstream region. 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 these alternative start codons than other strains, creating a hypothesis that differences in transcript length could dramatically alter phenotype[369]. To determine if *w*Rec and *w*Suzi *wmk* transgenes would be similarly sensitive to transcript changes, we also expressed these two homologs with the same alternative start codon previously tested (9 amino acids upstream). As expected, these two homologs also lost their phenotype of killing all hosts with only the 9 amino acids added to the 5' end of the gene (Fig. 6-4a). Instead, all expressing flies were able to survive. Returning to the RNA secondary structures, we find that simply adding the nucleotides at the 5' end of each homolog resulted in significant differences in RNA secondary structure for each transgene, with new loops forming or the position of the start codon in the overall structure substantially shifting (Figs. 6-4b and 6-4c). Thus, a small amino acid change can lead to large changes in predicted RNA structure, and altered phenotypes in *wmk* homologs, in line with previous findings that small transcriptional-level changes substantially affect phenotype.

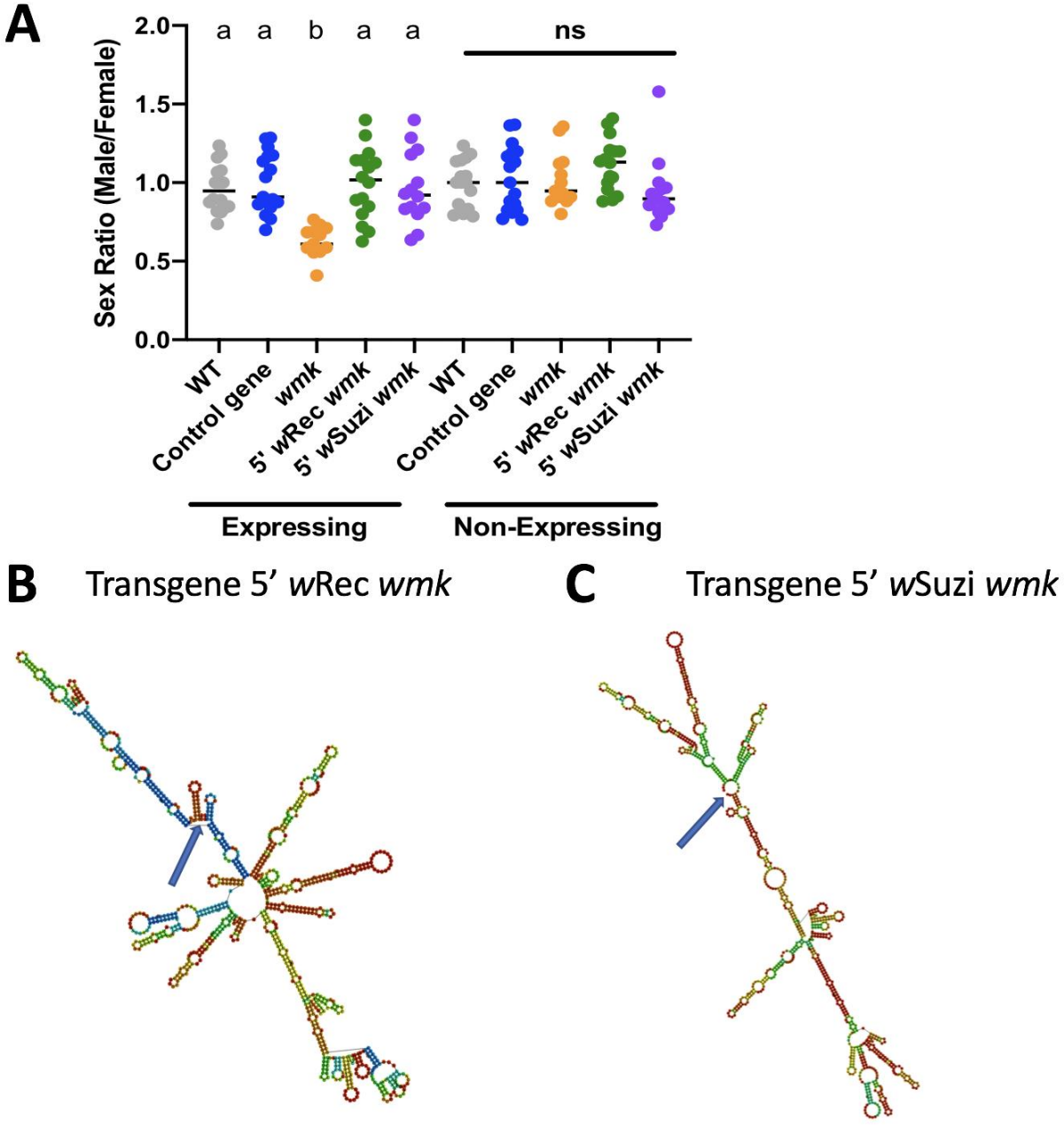


Figure 6-4. *w*Rec and *w*Suzi transgenes expressed with an alternative start codon lose their transgenic phenotypes.

(a) Sex ratios of adult flies either expressing (*Act5c*-Gal4) or not expressing (CyO) the indicated genes. Each sample point represents the adult offspring 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. (b & c) RNA secondary structures of the *w*Rec and *w*Suzi *wmk* homologs with the alternative 5' start codon. Blue arrows point to the start codon of each gene.

Finally, in an attempt to identify which exact nucleotides contributed to the vast differences in phenotype among the homologs, we aligned their sequences and looked for any sequences 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 has demonstrated that changes at the 5' end of this gene affect phenotype[369] and since the first ten or so codons in model prokaryote genes are known to substantially affect mRNA structure and resulting translation rate[479], we focused on the earlier codon. This codon, which codes for serine in all homologs, segregates among sequences by phenotype. The HA tag and *w*Mel strains have a TCG codon, *w*Suzi has TCC, and *w*Rec has AGC (Fig. 6-5a). To functionally test if this codon alone accounted for phenotype differences, we created three strains: 1) *w*Mel *wmk* created again with no changes, 2) *w*Mel *wmk* with the codon changed to the AGC from the homolog in *w*Rec, and 3) *w*Mel *wmk* with the codon changed to the TCC from the homolog in *w*Suzi.

When these three genes were transgenically expressed, the first strain still caused a biased sex ratio, but the single codon change for the second and third strains ablated any phenotype and resulted in a non-biased sex ratio and a normal number of expressing flies (ie, no all-killing phenotype) (Fig. 6-5b). One single codon change, with no corresponding amino acid change, was sufficient to alter the sex ratio phenotype, although the codon changes alone did not recapitulate the all-killing phenotypes of their corresponding homologs. Comparing the RNA secondary structures does demonstrate some similarities and differences among the transcripts (Fig. 6-3b). Indeed, the RNA secondary structure is significantly different in each genotype based on a single codon change (Figs. 6-5c and 6-5d).

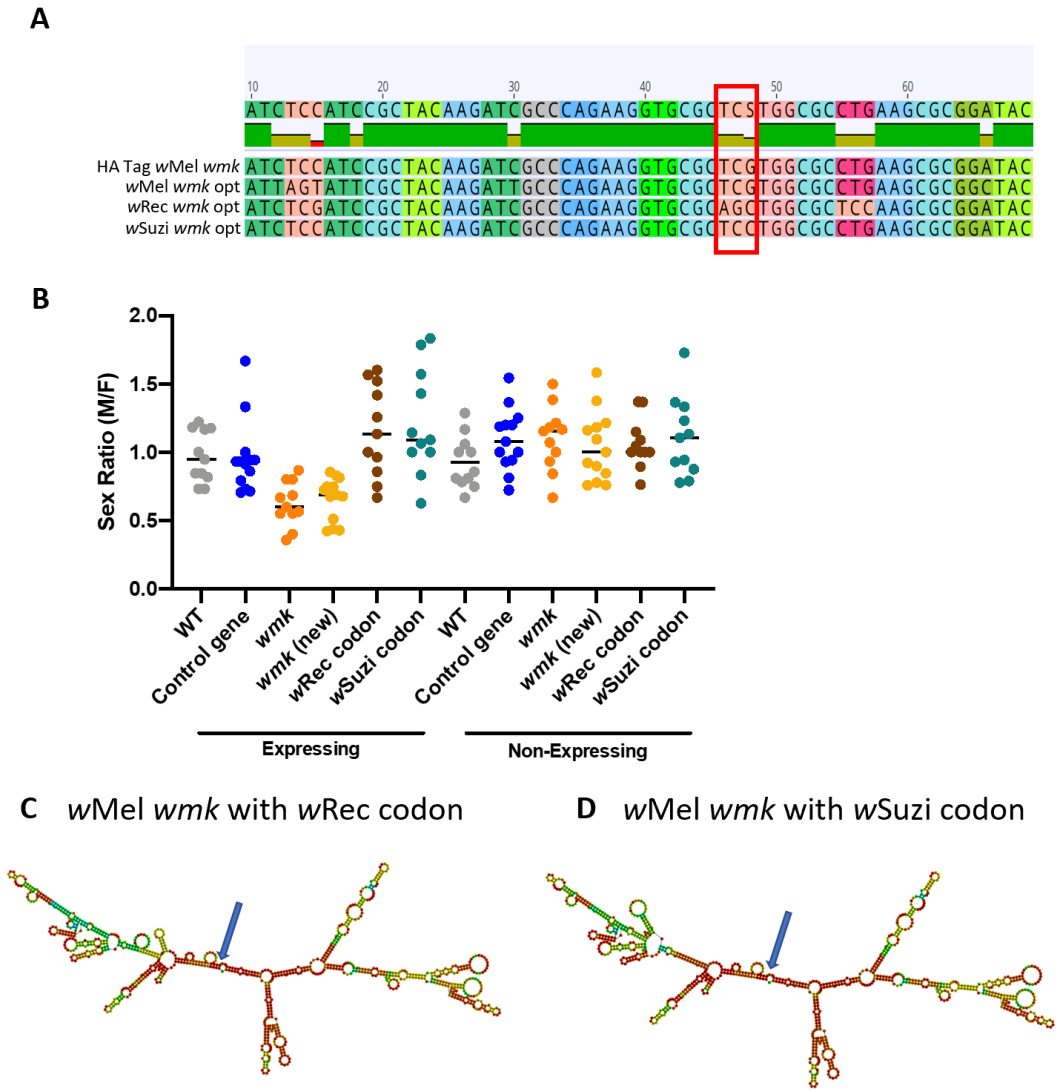


Figure 6-5. Changing the codon, but not the encoded amino acid, of the thirteenth amino acid position in *wmk* significantly alters resulting phenotype.
(a) Sequence alignment of the indicated genotypes. The red box outlines where the genotypes cluster by phenotype. The "*w*Mel *wmk* opt" and "HA tag *w*Mel *wmk*" genotypes share the same codon in this position and both induce male-specific death. The "*w*Rec *wmk* opt" and "*w*Suzi *wmk* opt" genotypes both exhibit different codons from the previous two and exhibit an all-killing phenotype. (b) Sex ratios of adult flies either expressing (*Act5c*-Gal4) or not expressing (CyO) the indicated genes. Each sample point represents the adult offspring 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. (c & d) RNA secondary structures of the *w*Mel *wmk* homolog with either the "*w*Rec *wmk* opt" or "*w*Suzi *wmk* opt" codon in the thirteenth amino acid codon position.

Discussion

There are several key findings from these experiments regarding *wmk* and its putative role in *Wolbachia*-mediated male killing, as well as broader implications of the work. First, the results neither confirm nor refute the role of *wmk* in *Wolbachia* male killing. If the *w*Rec male-killing homolog had recapitulated the phenotype as well, it would have provided support, but not proof of *wmk* as a male-killer. No recapitulation also does not disprove *wmk* as a male-killer, since the *w*Rec does have an amino acid difference that could be critical, and there is an extra layer of artificiality in that *w*Rec does not infect *D. melanogaster*, while *w*Mel (the strain the original gene is from) does. The difference in phenotype by similar homologs and the large phenotypic changes that result from small sequence changes do, however, complicate the scenario. Importantly, previous work demonstrated that not only did *wmk* induce a sex ratio bias, but it also is only one of seven genes shared across all sequenced male-killers, it recapitulated embryonic defects typical of natural infection, it is naturally expressed in embryos, and it induces DNA damage in association with dosage compensation activity, which is also typical of infection[160]. These previous results would be very unlikely if *wmk* did not have a role in male killing. The results here do not disprove *wmk*, but if *wmk* is a true male-killing gene, it suggests that there is a more complicated relationship between genotype and phenotype than presence and absence of the gene, or even host genetic background. Related, though it does not clarify the role of *wmk* in male killing, it does introduce a hypothesis that the partial sex ratio bias induced by *wmk* may in fact simply be due to the limits of the transgenic expression system. If small changes in the gene lead to large phenotype shifts, then it is plausible that the partial sex ratio bias is simply due to insufficient expression or misfolding of the RNA in the transgenic context that leads to a partial phenotype. Indeed, past and current results suggest that it may be pure luck that the original *wmk* transgene gave a male-killing-like phenotype at all and that testing had not started with a transgene resulting in a different phenotype. Confirmation or refutation of the role of *wmk* in natural male killing remains to be assessed in future experiments, which should be done in natural contexts.

The results also expand on previous findings regarding the complex relationship between *wmk* transcript length and resulting phenotype. Previous work demonstrated a correlation between longer transcripts and lack of male killing which was supported by functional tests of a transgene with an alternative start codon that resulted in ablation of the sex ratio bias phenotype[369]. We previously hypothesized that transcriptional level differences between strains may account for

phenotype differences. The current work expands on this by demonstrating that it is not only the length, but codon identity internal to the gene that can alter the phenotype. Importantly, this indicates that transgenic assessment of male killing in a non-native host with artificial expression will be difficult moving forward. Indeed, it is standard practice to codon-optimize genes for maximized host expression when testing endosymbiont gene function[158,160,248], however, this work demonstrates that this practice may distort results in some cases. Importantly, the algorithms that are used to codon-optimize are regularly updated. This means that strains created at different times may have significantly different genotypes. Indeed, the original wmk transgene was created several years before the updated one (Fig. 6-5) and the other transgenes tested here, under an older algorithm. The original and updated wmk transgenes have different sequences due to an updated codon-optimization algorithm. In the future, caution will be warranted in interpretation of transgenic results, which is especially true in experiments testing genes in non-native contexts. In addition, it may become more critical to compare transgenic results across strains that were created using the same codon-optimization algorithm.

On a broader level, it is informative that distantly-related homologs result in no phenotype (Fig. 6-1), but more similar homologs do result in phenotypes (Fig. 6-2), even if they are complex. This is somewhat in line with expectations relating to natural male-killing strains. Indeed, when *Wolbachia* are transferred between similar host species, they can often induce male killing in the similar, new host[266,267,341]. But, attempts to transfer male-killers to more distant hosts have resulted in loss of phenotype[342]. This suggests that male-killers need to be relatively closely-adapted to their hosts to induce a phenotype, where the toxin presumably must match the host target closely but the target becomes more different with phylogenetic distance between hosts. Thus, on a broad level, the results that distant homologs have no phenotype and similar homologs do impact the host in complex ways do somewhat align with expectations of male killing.

In addition, whether or not *wmk* is involved in natural male killing, this work has displayed an important principle that transgenic testing of *Wolbachia* genes has limits. Moving forward, it will likely become more difficult to study reproductive parasitism in unnatural contexts, and it may simply be luck that any result was found with *wmk*. Although transgenic testing may provide a suitable starting point in functional analyses of endosymbiont genes, researchers should strive to test hypotheses in more natural contexts. This is for several reasons, but also because here we have shown that transgenics result in complex phenotypes that are difficult to interpret. It is possible

that the only cases of clear transgenic results will be those that are native infections in genetically-tractable hosts, such as *Spiroplasma poulsonii* male killing and *Wolbachia* cytoplasmic incompatibility in *D. melanogaster*[158,159,248].

Finally, reproductive parasitism and *Wolbachia* aside, it is an exciting finding on its own that a synonymous codon change may result in a strong change in phenotype. This may inform future work on the expression and function of prophage genes in eukaryotes and provides a basis for future study of this phenomenon.

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[160], 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[160]. Briefly, twenty 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 set up in vials with CMY media. 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). Any vials with fewer than 50 adult offspring were removed from the analysis, as this indicates either poor egg laying or abnormally low egg hatching. The number of adult offspring per vial ranges from 50-170, with a mean of about 120.

RNA secondary structures

RNA secondary structures were generated by uploading the nucleotide sequences of the indicated gene RNA fold web server[489,490]. The structures shown are the graphical outputs of the MFE (minimum free energy) secondary structures.

Gene expression

Gene expression was measured in *Drosophila* embryos aged 4-5 h AED. Each point represents the RNA of 30 pooled embryos from crosses between 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 collection chambers. A grape juice agar plate with yeast was placed in each bottle. 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 the same bottle) 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 D-1. 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).

Sequence alignment

The gene sequence alignment of different *wmk* transgenes was done in Genious Pro v.2019.2. Codons are colored by amino acid and the height and color of the bar underneath the consensus sequence indicates degree of similarity across all sequences.

Statistical analyses

All statistical analyses for sex ratios and hatch rates were performed using GraphPad Prism 8 software. For sex ratios, a non-parametric Kruskal-Wallis one-way ANOVA followed by Dunn's test of multiple corrections was applied to all gene-expressing categories, followed by the same test but on all non-expressing categories.

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.

Chapter 7

Conclusions and future directions

Hubert Simmonds and others originally recorded female-biased sex ratios in butterfly populations over the last century, and since then the field has made significant advancements in our understanding of male killing and evolution as well as its genetic and molecular underpinnings. However, many questions remain that will be the subject of future investigations.

Are There More Male-Killers in Nature?

There are likely more male-killing taxa in nature that remain undiscovered. Given that so many diverse taxa have been discovered without any comprehensive search efforts (Table 2-1), the diversity of arthropods as a whole (85% of animal species), and the very recent findings of male-killing viruses[28,125], it is likely that the list will continue to expand. Especially as many of the known hosts are biased by their pest status, geography, or ease of rearing, etc., it is likely that many more male-killers will be discovered in the future. The list of affected host and causative microbial taxa will undoubtedly expand with time given the apparent ease of evolving male killing. This list will include additional cases of known microbial or host taxa with new hosts or microbes, respectively, as well as completely new symbioses. Future work should focus on screening additional, diverse arthropods not only for known microbes, but also for new ones, as they may be easily overlooked. In particular, the discoveries of viral male-killers are relatively new and the viruses are not characterized, and there may be many others that can cause the phenotype.

Why is Male Killing So Taxonomically Diverse Compared to Other Manipulations?

Compared to other reproductive manipulations, male killing occurs in a surprisingly wide diversity of hosts spanning many major insect orders and arachnids and by a wide variety of microbes including bacteria, viruses, and fungi. Parthenogenesis and feminization occur by fewer microorganisms (*Wolbachia*, *Cardinium*, and *Rickettsia* for parthenogenesis and *Wolbachia*, *Cardinium*, and microsporidia for feminization) and in a narrower range of hosts (hymenopterans and some arachnids for parthenogenesis and five major insect orders and amphipods for feminization)[130]. CI is only known to occur by *Wolbachia* or *Cardinium* but in a wide variety of hosts[23]. This is the basis for male killing's label as an "easily-evolved" phenotype[130,258]. Of these

phenotypes, male-killing may seem the least likely to evolve due to the delicate ecology that would be required to maintain it over time, but it remains common by some measures. Based on genetics, it has evolved at least twice, and most likely more times, independently. What makes male killing such a successful and broadly-employed strategy? Although many reasons have been proposed and tested for the advantages of male killing (Chapter 2), it is still unclear, and the reasons underlying its surprising success should continue to be researched. The maintenance of male killing over time in so many hosts by so many microbes remains one of the biggest evolutionary mysteries of the phenotype. In addition, since we know there are multiple cases of independent evolution, a question that remains is how did male-killing evolve? There must be some common factor(s) linking the cases together. What are all of the genetic and ecological factors that, combined, lead to development of male killing over and over in nature?

Despite the "ease" of evolution of male killing and its diversity, many host populations maintain male-killers at widely-varying proportions. For example, *w*Bif-infected *D. bifasciata* are found at a rate of 0-7% in wild populations[325], while *Wolbachia*-infected *A. encedon* butterflies are often at a near complete prevalence in wild populations[491]. What are the factors that determine the prevalence of a male-killer, how can it be maintained at such low levels in some cases while not wiping out the host population at high levels in others? Some have suggested that factors like incomplete transmission may play a role in the latter, but what determines complete or incomplete transmission in some cases vs others? This is a question that has been asked previously, but some aspects of it remain unclear and should be further researched[491].

How Do the Male-Killer and Host Evolve Over Time?

Related to the above questions of maintenance over time, how does a male-killing symbiosis evolve over time? Some have suggested extinction, and some have suggested that hosts will evolve new sex determination systems in an arms race with the bacteria[346]. In at least the case of *Wolbachia*-infected *Ostrinia* moths, the microbe has evolved a feminizing trait that accounts for degradation of feminizing factors in the host in an apparent mutualism[365]. As stated before, it takes a balance of ecological factors to maintain male killing, but at some point, the balance will change. What are all of the possible outcomes for host and microbe over long periods of evolutionary time? How do these changes occur on a genetic and molecular level?

Are There More Male-Killing Genes in Nature?

Just as there are likely more male-killers in nature, it is likely we have not discovered all male-killing genes. Based on the diversity of affected hosts and microbial taxa, the different sex determination systems of the arthropods, and the varied phenotypes and strategies employed by different male-killers, it is probable that more genes have evolved to induce male killing (Chapter 2, Table 2-1). Further, *Sp*AID does not exist outside of *Spiroplasma* and *wmk* is unknown outside of *Wolbachia* (and apparently arose within the clade)[160,369], so it is unclear how other male-killers such as *Rickettsia* and *Arsenophonus* kill males. Even within a microbial taxon, it is possible that there are additional genes. For example, *wmk* may only kill males of certain *Drosophila* species, but may not kill lepidopterans that exhibit differences in their phenotypes (Chapter 2). There may also be modifier genes that work alongside *Sp*AID or *wmk* in natural contexts. There are undoubtedly many more genetic discoveries to be made in male killing, which will need to be the focus of future investigations.

In addition, it is unclear why (at least for *Wolbachia*) other phenotypes such as CI have an apparent connection to male killing. When male killing is hidden (or there is phenotype switching between hosts), CI or other phenotypes often appear (Table 2-2), and it is unclear what the basis of the connection may be. It is possible that it is simply ecological in that having two phenotypes available gives a *Wolbachia* strain a fitness advantage. Transgenic testing did not support a functional role of *wmk*, *cifA*, and many other *Wolbachia* genes in reproductive parasitism, but there may be more genetic connections that remain to be uncovered[369]. In addition, the basis of phenotype switching on the host side remains a mystery. Any genetic connections between commonly coincident phenotypes remain open questions.

As a caution, the field will need to be careful when studying these phenotypes in strains that have been reared in the lab for long periods of time. If phenotype-switching is common in nature and ecologically-relevant, then lab conditions may relax selection on one phenotype or another, and the genes may be lost. Indeed, *w*Bif was reported to cause CI two decades ago[262], but recent sequencing on a strain reared in the lab during that time revealed that it lacks the *cifB* gene required to induce CI[160]. This may be a case where a parasitism gene was lost due to selection for male killing in the lab. It remains to be tested if *w*Bif is still capable of inducing CI, but transgenic experiments in *D. melanogaster* suggest it is unlikely, especially as *cifA* is the putative rescue factor and *cifB* is required for CI induction[158]. Thus, any future investigations into the genetics of

male killing or CI should proceed with caution, as lab-reared strains may not reflect natural genetics. In addition, lack of a detectable male-killing phenotype should not be taken as a sign that the strain cannot cause male killing, since hidden male killing and phenotype switching mean that the phenotype could occur in other conditions. Thus, it will be difficult to demonstrate a true negative strain for male killing, and possibly CI as well.

What are the Mechanisms of Male Killing and How Does Resistance Arise?

So far, the field has uncovered many differences in the mechanisms of male killing from one host-microbe pair to another. In some cases, the mechanism is interference with dosage compensation[359,360], others may involve dosage compensation[357], some microbes kill males through lethal feminization[364], while still others interfere with masculinization[366]. It therefore appears that there may be many different ways to kill males, even within a particular sex determination system or host species. There are likely other, undiscovered mechanisms, so continued research on this will be important. Importantly, although we have a general idea of mechanism, especially with *Spiroplasma*, we do not have the full picture. It will be important to keep identifying host targets and to make progress on our molecular understanding of male-killing mechanism. This will be difficult moving forward for most cases since, except for *Spiroplasma poulsonii* in *D. melanogaster*, male killing does not occur naturally in genetically tractable species. It is possible that progress will be slow and technically difficult in this area until developments are made in the genetic manipulation of either other hosts with natural male-killing infections, or the microbes themselves. This will be particularly critical in assessment of *wmk* as a *Wolbachia* male-killing candidate.

Although many hosts are known to resist male killing (Table 2-2), others exhibit no ability to resist the phenotype across entire populations[264,330]. Theoretically, resistance is due to mutations in the host target gene that change it so that the male-killing toxin can no longer function. Although no specific host targets have been named thus far, there are advancements in that area[331]. Once particular genes are identified, this will enable co-evolution studies to test arms race hypotheses and better understand male-killing dynamics in a host population. It will also advance studies of mechanism, and allow prediction of which species will be susceptible to male killing, and which ones are likely to be able to resist it. Currently, it is unclear why some populations commonly develop resistance such as *H. bolina*[328], while others do not, such as *D. innubila*[330].

What is the Level of Specificity Between a Male-Killer and Its Host?

Related to the mechanism and resistance, how specific is a pairing between a particular male-killer and its host? Transinfection experiments demonstrate a clear trend that male killing may occur in non-native hosts, but typically only if they are closely-related, unlike CI which may occur through one strain in distantly-related species[368]. Indeed, transfer of male-killing *w*Inn from *D. innubila* to distant hosts *D. melanogaster* and *D. simulans* results in loss of phenotype[342]. In addition, experiments testing *wmk* homologs from phylogenetically distant hosts in *D. melanogaster* fail to uncover a phenotype (Chapter 6), suggesting that there is a high level of co-adaptation required between host and microbe for male killing to occur. What this specificity indicates is that whatever molecules are targeted in these hosts are likely not well-conserved. This is as opposed to CI, which likely targets well-conserved cellular components given the ability of CI strains to induce the phenotype in phylogenetically distant hosts. The inability of male killing to be induced in unrelated species by the same strain also supports the hypothesis of multiple mechanisms in nature, as the toxin would need to adapt to male-specific molecules in each particular host, which differ greatly in each host species. It also supports the finding that resistance is common, as high levels of specificity suggest that even minor differences in host target can negate the phenotype. On a practical level, high specificity may be useful in pest or vector control to keep the male-killer from spreading beyond a target population, but would also require development or discovery of a suitable toxin for each host, and there may be a risk of resistance.

How Do Male-Killing Symbionts Interact with the Rest of the Microbiome?

Returning to the topics introduced in Chapter 1, there is significant progress in understanding how male-killers specifically interact with their hosts and how they impact host evolution. However, some questions remain unexplored in male-killing. In particular, what interactions are there between male-killers and the rest of the microbiome? Do other microbes play a role in any of the unanswered questions posed here? For example, do other microbes impact transmission of male-killers, and therefore do they mediate maintenance of male-killers in a population in any way? Expanding beyond the single host-microbe pair in male-killing symbioses will be an exciting future direction for the field.

References

1 Oulhen, N., Schulz, B. J. & Carrier, T. J. English translation of Heinrich Anton de Bary's 1878 speech, 'Die Erscheinung der Symbiose' ('De la symbiose'). *Symbiosis* **69**, 131-139, doi:10.1007/s13199-016-0409-8 (2016).

2 De Bary, A. Die Erscheinung der Symbiose. Vortrag auf der Versammlung der Naturforscher und Aertze zu Cassel. *Trubner, Strasbourg*, 21-22 (1879).

3 Pierantoni, U. L'origine di alcuni organi d'Icerya purchasi e la simbiosi ereditaria. *Boll. Soc. Nat. Napoli* **23**, 147-150 (1909).

4 Sapp, J. Paul Buchner (1886-1978) and hereditary symbiosis in insects. *Int Microbiol* **5**, 145-150 (2002).

5 Funkhouser, L. J. & Bordenstein, S. R. Mom knows best: the universality of maternal microbial transmission. *PLoS Biol* **11**, e1001631, doi:10.1371/journal.pbio.1001631 (2013).

6 Werren, J. H., Baldo, L. & Clark, M. E. Wolbachia: master manipulators of invertebrate biology. *Nat Rev Microbiol* **6**, 741-751, doi:10.1038/nrmicro1969 (2008).

7 Chen, X., Li, S. & Aksoy, S. Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus Glossina and its bacteriome-associated endosymbiont, Wigglesworthia glossinidia. *Journal of molecular evolution* **48**, 49-58 (1999).

8 Douglas, A. E. Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria Buchnera. *Annual Review of Entomology* **43**, 17-37, doi:10.1146/annurev.ento.43.1.17 (1998).

9 Segata, N. *et al.* The reproductive tracts of two malaria vectors are populated by a core microbiome and by gender-and swarm-enriched microbial biomarkers. *Scientific reports* **6**, 24207 (2016).

10 Oliver, K. M., Campos, J., Moran, N. A. & Hunter, M. S. Population dynamics of defensive symbionts in aphids. *Proceedings of the Royal Society of London B: Biological Sciences* **275**, 293-299 (2008).

11 Jia, D. *et al.* Insect symbiotic bacteria harbour viral pathogens for transovarial transmission. *Nature microbiology* **2**, 17025 (2017).

12 Toh, H. *et al.* Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of Sodalis glossinidius in the tsetse host. *Genome research* **16**, 149-156 (2006).

13 Sabree, Z. L., Kambhampati, S. & Moran, N. A. Nitrogen recycling and nutritional provisioning by Blattabacterium, the cockroach endosymbiont. *Proceedings of the National Academy of Sciences* **106**, 19521-19526 (2009).

14 Moran, N. A., McCutcheon, J. P. & Nakabachi, A. Genomics and evolution of heritable bacterial symbionts. *Annual review of genetics* **42**, 165-190 (2008).

15 Gibson, C. M. & Hunter, M. S. Extraordinarily widespread and fantastically complex: comparative biology of endosymbiotic bacterial and fungal mutualists of insects. *Ecology Letters* **13**, 223-234 (2010).

16 Knell, R. J. & Webberley, K. M. Sexually transmitted diseases of insects: distribution, evolution, ecology and host behaviour. *Biological reviews of the Cambridge Philosophical Society* **79**, 557-581 (2004).

17 Cooley, J. R., Marshall, D. C. & Hill, K. B. A specialized fungal parasite (Massospora cicadina) hijacks the sexual signals of periodical cicadas (Hemiptera: Cicadidae: Magicicada). *Scientific reports* **8**, 1432 (2018).
18 Gibson, C. M. & Hunter, M. S. Inherited fungal and bacterial endosymbionts of a parasitic wasp and its cockroach host. *Microbial ecology* **57**, 542 (2009).
19 Boldbaatar, D. *et al.* Tick vitellogenin receptor reveals critical role in oocyte development and transovarial transmission of Babesia parasite. *Biochemistry and cell biology* **86**, 331-344 (2008).
20 Bordenstein, S. R. & Bordenstein, S. R. Eukaryotic association module in phage WO genomes from Wolbachia. *Nature communications* **7**, 13155, doi:10.1038/ncomms13155 (2016).
21 van der Wilk, F., Dullemans, A. M., Verbeek, M. & van den Heuvel, J. F. Isolation and characterization of APSE-1, a bacteriophage infecting the secondary endosymbiont of Acyrthosiphon pisum. *Virology* **262**, 104-113 (1999).
22 Wei, J. *et al.* Vector development and vitellogenin determine the transovarial transmission of begomoviruses. *Proceedings of the National Academy of Sciences* **114**, 6746-6751 (2017).
23 Hurst, G. D. & Frost, C. L. Reproductive parasitism: maternally inherited symbionts in a biparental world. *Cold Spring Harbor perspectives in biology* **7**, a017699 (2015).
24 Chevalier, F. *et al.* Feminizing Wolbachia: a transcriptomics approach with insights on the immune response genes in Armadillidium vulgare. *BMC microbiology* **12**, S1 (2012).
25 Haine, E. R., Motreuil, S. & Rigaud, T. Infection by a vertically-transmitted microsporidian parasite is associated with a female-biased sex ratio and survival advantage in the amphipod Gammarus roeseli. *Parasitology* **134**, 1363-1367 (2007).
26 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).
27 Gotoh, T., Noda, H. & Ito, S. Cardinium symbionts cause cytoplasmic incompatibility in spider mites. *Heredity* **98**, 13-20, doi:10.1038/sj.hdy.6800881 (2007).
28 Kageyama, D., Yoshimura, K., Sugimoto, T. N., Katoh, T. K. & Watada, M. Maternally transmitted non-bacterial male killer in Drosophila biauraria. *Biol Lett* **13**, doi:10.1098/rsbl.2017.0476 (2017).
29 Ironside, J. E. *et al.* Two species of feminizing microsporidian parasite coexist in populations of Gammarus duebeni. *Journal of evolutionary biology* **16**, 467-473 (2003).
30 Bandi, C., Dunn, A. M., Hurst, G. D. & Rigaud, T. Inherited microorganisms, sex-specific virulence and reproductive parasitism. *Trends in parasitology* **17**, 88-94 (2001).
31 Perilla-Henao, L. M. & Casteel, C. L. Vector-borne bacterial plant pathogens: interactions with hemipteran insects and plants. *Frontiers in Plant Science* **7**, 1163 (2016).
32 Gray, S. M. & Banerjee, N. Mechanisms of arthropod transmission of plant and animal viruses. *Microbiol. Mol. Biol. Rev.* **63**, 128-148 (1999).
33 Hogenhout, S. A., Ammar, E.-D., Whitfield, A. E. & Redinbaugh, M. G. Insect vector interactions with persistently transmitted viruses. *Annu. Rev. Phytopathol.* **46**, 327-359 (2008).
34 Baumann, P. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu. Rev. Microbiol.* **59**, 155-189 (2005).

35 Xie, J., Butler, S., Sanchez, G. & Mateos, M. Male killing Spiroplasma protects Drosophila melanogaster against two parasitoid wasps. *Heredity* **112**, 399 (2014).
36 Brownlie, J. C. & Johnson, K. N. Symbiont-mediated protection in insect hosts. *Trends in microbiology* **17**, 348-354 (2009).
37 Degnan, P. H. & Moran, N. A. Evolutionary genetics of a defensive facultative symbiont of insects: exchange of toxin-encoding bacteriophage. *Molecular Ecology* **17**, 916-929 (2008).
38 Kellner, R. L. Molecular identification of an endosymbiotic bacterium associated with pederin biosynthesis in Paederus sabaeus (Coleoptera: Staphylinidae). *Insect biochemistry and molecular biology* **32**, 389-395 (2002).
39 Feldhaar, H. *et al.* Nutritional upgrading for omnivorous carpenter ants by the endosymbiont Blochmannia. *BMC biology* **5**, 48 (2007).
40 Rock, D. I. *et al.* Context-dependent vertical transmission shapes strong endosymbiont community structure in the pea aphid, Acyrthosiphon pisum. *Molecular ecology* **27**, 2039-2056 (2018).
41 Zhu, L.-Y. *et al.* Wolbachia strengthens Cardinium-induced cytoplasmic incompatibility in the spider mite Tetranychus piercei McGregor. *Current microbiology* **65**, 516-523 (2012).
42 White, J. A., Kelly, S. E., Perlman, S. J. & Hunter, M. S. Cytoplasmic incompatibility in the parasitic wasp Encarsia inaron: disentangling the roles of Cardinium and Wolbachia symbionts. *Heredity* **102**, 483 (2009).
43 Hughes, G. L. *et al.* Native microbiome impedes vertical transmission of Wolbachia in Anopheles mosquitoes. *Proceedings of the National Academy of Sciences* **111**, 12498-12503 (2014).
44 Bellinvia, S., Johnston, P. R., Reinhardt, K. & Otti, O. Bacterial communities of the reproductive organs of virgin and mated common bedbugs, Cimex lectularius. *Ecological Entomology* (2019).
45 Kondo, N., Shimada, M. & Fukatsu, T. Infection density of Wolbachia endosymbiont affected by co-infection and host genotype. *Biology letters* **1**, 488-491 (2005).
46 Skaljac, M., Zanic, K., Ban, S. G., Kontsedalov, S. & Ghanim, M. Co-infection and localization of secondary symbionts in two whitefly species. *BMC microbiology* **10**, 142 (2010).
47 Toju, H. & Fukatsu, T. Diversity and infection prevalence of endosymbionts in natural populations of the chestnut weevil: relevance of local climate and host plants. *Molecular Ecology* **20**, 853-868 (2011).
48 Tokura, M., Ohkuma, M. & Kudo, T. Molecular phylogeny of methanogens associated with flagellated protists in the gut and with the gut epithelium of termites. *FEMS microbiology ecology* **33**, 233-240 (2000).
49 Nimrat, S., Bart, A. N., Keatsaksit, A. & Vuthiphandchai, V. Microbial flora of spermatophores from black tiger shrimp (Penaeus monodon) declines over long-term cryostorage. *Aquaculture* **274**, 247-253 (2008).
50 Benhalima, K. & Moriyasu, M. Prevalence of bacteria in the spermathecae of female snow crab, Chionoecetes opilio (Brachyura: Majidae). *Hydrobiologia* **449**, 261-266 (2001).
51 Duron, O., Hurst, G. D., Hornett, E. A., Josling, J. A. & Engelstädter, J. High incidence of the maternally inherited bacterium Cardinium in spiders. *Molecular Ecology* **17**, 1427-1437 (2008).

52 Vala, F., Egas, M., Breeuwer, J. & Sabelis, M. Wolbachia affects oviposition and mating behaviour of its spider mite host. *Journal of evolutionary biology* **17**, 692-700 (2004).
53 Martin, O. Y. & Goodacre, S. L. Widespread infections by the bacterial endosymbiont Cardinium in arachnids. *The Journal of Arachnology* **37**, 106-109 (2009).
54 Vanthournout, B., Vandomme, V. & Hendrickx, F. Sex ratio bias caused by endosymbiont infection in the dwarf spider Oedothorax retusus. *Journal of arachnology*, 24-33 (2014).
55 Gaci, N., Borrel, G., Tottey, W., O'Toole, P. W. & Brugère, J.-F. Archaea and the human gut: new beginning of an old story. *World journal of gastroenterology: WJG* **20**, 16062 (2014).
56 Gilbert, J. A. *et al.* Current understanding of the human microbiome. *Nat Med* **24**, 392-400, doi:10.1038/nm.4517 (2018).
57 Turner, T. R., James, E. K. & Poole, P. S. The plant microbiome. *Genome biology* **14**, 209 (2013).
58 Bright, M. & Bulgheresi, S. A complex journey: transmission of microbial symbionts. *Nature Reviews Microbiology* **8**, 218, doi:10.1038/nrmicro2262 https://www.nature.com/articles/nrmicro2262#supplementary-information (2010).
59 Newton, I. L., Savytskyy, O. & Sheehan, K. B. Wolbachia utilize host actin for efficient maternal transmission in Drosophila melanogaster. *PLoS pathogens* **11**, e1004798 (2015).
60 Brumin, M., Levy, M. & Ghanim, M. Transovarial transmission of Rickettsia spp. and organ-specific infection of the whitefly Bemisia tabaci. *Applied and environmental microbiology* **78**, 5565-5574, doi:10.1128/aem.01184-12 (2012).
61 Herren, J. K., Paredes, J. C., Schüpfer, F. & Lemaitre, B. Vertical transmission of a Drosophila endosymbiont via cooption of the yolk transport and internalization machinery. *mBio* **4**, e00532-00512 (2013).
62 Wilkinson, T., Fukatsu, T. & Ishikawa, H. Transmission of symbiotic bacteria Buchnera to parthenogenetic embryos in the aphid Acyrthosiphon pisum (Hemiptera: Aphidoidea). *Arthropod structure & development* **32**, 241-245 (2003).
63 Dykstra, H. R. *et al.* Factors limiting the spread of the protective symbiont Hamiltonella defensa in Aphis craccivora aphids. *Appl. Environ. Microbiol.* **80**, 5818-5827 (2014).
64 Cheng, D. & Hou, R. Histological observations on transovarial transmission of a yeast-like symbiote in Nilaparvata lugens Stal (Homoptera, Delphacidae). *Tissue and Cell* **33**, 273-279 (2001).
65 Solter, L. F. Transmission as a predictor of ecological host specificity with a focus on vertical transmission of microsporidia. *Journal of invertebrate pathology* **92**, 132-140 (2006).
66 Huo, Y. *et al.* Transovarial transmission of a plant virus is mediated by vitellogenin of its insect vector. *PLoS pathogens* **10**, e1003949 (2014).
67 Fast, E. M. *et al.* Wolbachia enhance Drosophila stem cell proliferation and target the germline stem cell niche. *Science (New York, N.Y.)* **334**, 990-992, doi:10.1126/science.1209609 (2011).
68 Ferree, P. M. *et al.* Wolbachia utilizes host microtubules and Dynein for anterior localization in the Drosophila oocyte. *PLoS pathogens* **1**, e14, doi:10.1371/journal.ppat.0010014 (2005).
69 Koga, R., Meng, X.-Y., Tsuchida, T. & Fukatsu, T. Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte–embryo interface. *Proceedings of the National Academy of Sciences* **109**, E1230-E1237 (2012).

70 Salem, H., Florez, L., Gerardo, N. & Kaltenpoth, M. An out-of-body experience: the extracellular dimension for the transmission of mutualistic bacteria in insects. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20142957 (2015).
71 Kikuchi, Y. *et al.* Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. *Bmc Biology* **7**, 2 (2009).
72 Salem, H. *et al.* Drastic Genome Reduction in an Herbivore's Pectinolytic Symbiont. *Cell* **171**, 1520-1531.e1513, doi:10.1016/j.cell.2017.10.029 (2017).
73 Dominguez-Bello, M. G. *et al.* Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 11971-11975, doi:10.1073/pnas.1002601107 (2010).
74 Kaiwa, N. *et al.* Symbiont-supplemented maternal investment underpinning host's ecological adaptation. *Current Biology* **24**, 2465-2470 (2014).
75 Shukla, S. P., Vogel, H., Heckel, D. G., Vilcinskas, A. & Kaltenpoth, M. Burying beetles regulate the microbiome of carcasses and use it to transmit a core microbiota to their offspring. *Molecular Ecology* **27**, 1980-1991, doi:10.1111/mec.14269 (2017).
76 Pais, R., Lohs, C., Wu, Y., Wang, J. & Aksoy, S. The obligate mutualist Wigglesworthia glossinidia influences reproduction, digestion, and immunity processes of its host, the tsetse fly. *Applied and environmental microbiology* **74**, 5965-5974, doi:10.1128/aem.00741-08 (2008).
77 Nadal-Jimenez, P. *et al.* Genetic manipulation allows in vivo tracking of the life cycle of the son-killer symbiont, Arsenophonus nasoniae, and reveals patterns of host invasion, tropism and pathology. *Environmental Microbiology* **21**, 3172-3182 (2019).
78 Gottlieb, Y. *et al.* The transmission efficiency of tomato yellow leaf curl virus by the whitefly Bemisia tabaci is correlated with the presence of a specific symbiotic bacterium species. *Journal of virology* **84**, 9310-9317 (2010).
79 van den Heuvel, J. F., Verbeek, M. & van der Wilk, F. Endosymbiotic bacteria associated with circulative transmission of potato leafroll virus by Myzus persicae. *Journal of General Virology* **75**, 2559-2565 (1994).
80 Touret, F., Guiguen, F. & Terzian, C. Wolbachia influences the maternal transmission of the gypsy endogenous retrovirus in Drosophila melanogaster. *mBio* **5**, e01529-01514 (2014).
81 Dutra, H. L. *et al.* Wolbachia Blocks Currently Circulating Zika Virus Isolates in Brazilian Aedes aegypti Mosquitoes. *Cell host & microbe* **19**, 771-774, doi:10.1016/j.chom.2016.04.021 (2016).
82 Hegde, S., Rasgon, J. L. & Hughes, G. L. The microbiome modulates arbovirus transmission in mosquitoes. *Current opinion in virology* **15**, 97-102 (2015).
83 Watanabe, K., Yukuhiro, F., Matsuura, Y., Fukatsu, T. & Noda, H. Intrasperm vertical symbiont transmission. *Proceedings of the National Academy of Sciences* **111**, 7433-7437 (2014).
84 Damiani, C. *et al.* Paternal transmission of symbiotic bacteria in malaria vectors. *Current Biology* **18**, R1087-R1088 (2008).
85 De Vooght, L., Caljon, G., Van Hees, J. & Van Den Abbeele, J. Paternal transmission of a secondary symbiont during mating in the viviparous tsetse fly. *Molecular biology and evolution* **32**, 1977-1980 (2015).

86 Longdon, B. & Jiggins, F. M. Vertically transmitted viral endosymbionts of insects: do sigma viruses walk alone? *Proceedings of the Royal Society B: Biological Sciences* **279**, 3889-3898 (2012).
87 Longdon, B. & Jiggins, F. M. Paternally transmitted parasites. *Current Biology* **20**, R695-R696 (2010).
88 Engelstädter, J. & Hurst, G. D. What use are male hosts? The dynamics of maternally inherited bacteria showing sexual transmission or male killing. *The American Naturalist* **173**, E159-E170 (2009).
89 Ironside, J. E., Smith, J. E., Hatcher, M. J. & Dunn, A. M. Should sex-ratio distorting parasites abandon horizontal transmission? *BMC evolutionary biology* **11**, 370 (2011).
90 Otti, O. Genitalia-associated microbes in insects. *Insect science* **22**, 325-339 (2015).
91 Bolling, B. G., Olea-Popelka, F. J., Eisen, L., Moore, C. G. & Blair, C. D. Transmission dynamics of an insect-specific flavivirus in a naturally infected Culex pipiens laboratory colony and effects of co-infection on vector competence for West Nile virus. *Virology* **427**, 90-97 (2012).
92 Moran, N. A. & Dunbar, H. E. Sexual acquisition of beneficial symbionts in aphids. *Proceedings of the National Academy of Sciences* **103**, 12803-12806 (2006).
93 Sookar, P., Bhagwant, S. & Allymamod, M. Effect of Metarhizium anisopliae on the fertility and fecundity of two species of fruit flies and horizontal transmission of mycotic infection. *Journal of insect science* **14**, 100 (2014).
94 Adamo, S. A., Kovalko, I., Easy, R. H. & Stoltz, D. A viral aphrodisiac in the cricket Gryllus texensis. *Journal of Experimental Biology* **217**, 1970-1976 (2014).
95 García-Munguía, A. M., Garza-Hernández, J. A., Rebollar-Tellez, E. A., Rodríguez-Pérez, M. A. & Reyes-Villanueva, F. Transmission of Beauveria bassiana from male to female Aedes aegypti mosquitoes. *Parasites & vectors* **4**, 24 (2011).
96 Lung, O., Kuo, L. & Wolfner, M. F. Drosophila males transfer antibacterial proteins from their accessory gland and ejaculatory duct to their mates. *J Insect Physiol* **47**, 617-622 (2001).
97 Zhong, W. *et al.* Immune anticipation of mating in Drosophila: Turandot M promotes immunity against sexually transmitted fungal infections. *Proceedings. Biological sciences* **280**, 20132018, doi:10.1098/rspb.2013.2018 (2013).
98 Esteves, E. *et al.* Antimicrobial activity in the tick Rhipicephalus (Boophilus) microplus eggs: Cellular localization and temporal expression of microplusin during oogenesis and embryogenesis. *Dev Comp Immunol* **33**, 913-919, doi:10.1016/j.dci.2009.02.009 (2009).
99 Kaya, M. *et al.* New chitin, chitosan, and O -carboxymethyl chitosan sources from resting eggs of Daphnia longispina(Crustacea); with physicochemical characterization, and antimicrobial and antioxidant activities. *Biotechnol Bioproc E* **19**, 58-69, doi:10.1007/s12257-013-0488-9 (2014).
100 Peng, Y., Grassl, J., Millar, A. H. & Baer, B. Seminal fluid of honeybees contains multiple mechanisms to combat infections of the sexually transmitted pathogen Nosema apis. *Proceedings of the Royal Society B: Biological Sciences* **283**, 20151785 (2016).
101 Gauthier, L. *et al.* Viruses associated with ovarian degeneration in Apis mellifera L. queens. *PLOS one* **6**, e16217 (2011).
102 Burand, J. P. The sexually transmitted insect virus, Hz-2V. *Virologica Sinica* **24**, 428 (2009).

103 Himler, A. G. *et al.* Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. *Science (New York, N.Y.)* **332**, 254-256 (2011).
104 Weeks, A. R., Turelli, M., Harcombe, W. R., Reynolds, K. T. & Hoffmann, A. A. From parasite to mutualist: rapid evolution of Wolbachia in natural populations of Drosophila. *PLoS biology* **5**, e114 (2007).
105 Patot, S., Lepetit, D., Charif, D., Varaldi, J. & Fleury, F. Molecular detection, penetrance, and transmission of an inherited virus responsible for behavioral manipulation of an insect parasitoid. *Appl. Environ. Microbiol.* **75**, 703-710 (2009).
106 Pannebakker, B. A., Loppin, B., Elemans, C. P., Humblot, L. & Vavre, F. Parasitic inhibition of cell death facilitates symbiosis. *Proceedings of the National Academy of Sciences* **104**, 213-215 (2007).
107 Zchori-Fein, E., Borad, C. & Harari, A. R. Oogenesis in the date stone beetle, Coccotrypes dactyliperda, depends on symbiotic bacteria. *Physiological Entomology* **31**, 164-169 (2006).
108 Snook, R. R., Cleland, S. Y., Wolfner, M. F. & Karr, T. L. Offsetting effects of Wolbachia infection and heat shock on sperm production in Drosophila simulans: analyses of fecundity, fertility and accessory gland proteins. *Genetics* **155**, 167-178 (2000).
109 Ebbert, M. A., Marlowe, J. L. & Burkholder, J. J. Protozoan and intracellular fungal gut endosymbionts in Drosophila: prevalence and fitness effects of single and dual infections. *Journal of invertebrate pathology* **83**, 37-45 (2003).
110 Gibson, C. M. & Hunter, M. S. Negative fitness consequences and transmission dynamics of a heritable fungal symbiont of a parasitic wasp. *Appl. Environ. Microbiol.* **75**, 3115-3119 (2009).
111 Yixin, H. Y., Woolfit, M., Rancès, E., O'Neill, S. L. & McGraw, E. A. Wolbachia-associated bacterial protection in the mosquito Aedes aegypti. *PLoS neglected tropical diseases* **7**, e2362 (2013).
112 Braquart-Varnier, C. *et al.* The mutualistic side of Wolbachia–isopod interactions: Wolbachia mediated protection against pathogenic intracellular bacteria. *Frontiers in microbiology* **6**, 1388 (2015).
113 Nikoh, N. *et al.* Genomic Insight into Symbiosis-Induced Insect Color Change by a Facultative Bacterial Endosymbiont, "Candidatus Rickettsiella viridis". *mBio* **9**, doi:10.1128/mBio.00890-18 (2018).
114 Funkhouser-Jones, L. J., van Opstal, E. J., Sharma, A. & Bordenstein, S. R. The Maternal Effect Gene Wds Controls Wolbachia Titer in Nasonia. *Current biology : CB* **28**, 1692-1702.e1696, doi:10.1016/j.cub.2018.04.010 (2018).
115 Moriyama, M. *et al.* Comparative transcriptomics of the bacteriome and the spermalege of the bedbug Cimex lectularius (Hemiptera: Cimicidae). *Applied entomology and zoology* **47**, 233-243 (2012).
116 Stutt, A. D. & Siva-Jothy, M. T. Traumatic insemination and sexual conflict in the bed bug Cimex lectularius. *Proceedings of the National Academy of Sciences* **98**, 5683-5687 (2001).
117 Reinhardt, K., Naylor, R. & Siva–Jothy, M. T. Reducing a cost of traumatic insemination: female bedbugs evolve a unique organ. *Proceedings of the Royal Society of London B: Biological Sciences* **270**, 2371-2375 (2003).
118 BÜCHNER, P. Studien an intracellularen Symbionten. IV-Die Bakteriensymbiose der Bettwanze. *Arch. Protistenk* **46**, 225-263 (1923).

119 Buchner, P. Endosymbiosis of animals with plant microorganisms. (1965).
120 Marshall, J. L. Rapid evolution of spermathecal duct length in the Allonemobius socius complex of crickets: species, population and Wolbachia effects. *PLoS One* **2**, e720 (2007).
121 Duron, O. *et al.* The diversity of reproductive parasites among arthropods: Wolbachia do not walk alone. *BMC Biol* **6**, 27, doi:10.1186/1741-7007-6-27 (2008).
122 Dunn, A. M. & Smith, J. E. Microsporidian life cycles and diversity: the relationship between virulence and transmission. *Microbes and Infection* **3**, 381-388 (2001).
123 Wang, F. *et al.* A novel negative-stranded RNA virus mediates sex ratio in its parasitoid host. *PLoS pathogens* **13**, e1006201 (2017).
124 Juchault, P., Louis, C., Martin, G. & Noulin, G. Masculinization of female isopods (Crustacea) correlated with non-Mendelian inheritance of cytoplasmic viruses. *Proceedings of the National Academy of Sciences* **88**, 10460-10464 (1991).
125 Nakanishi, K., Hoshino, M., Nakai, M. & Kunimi, Y. Novel RNA sequences associated with late male killing in Homona magnanima. *Proceedings of the Royal Society B: Biological Sciences* **275**, 1249-1254 (2008).
126 Dedeine, F., Bouletreau, M. & Vavre, F. Wolbachia requirement for oogenesis: occurrence within the genus Asobara (Hymenoptera, Braconidae) and evidence for intraspecific variation in A. tabida. *Heredity* **95**, 394 (2005).
127 Leclercq, S. *et al.* Birth of a W sex chromosome by horizontal transfer of Wolbachia bacterial symbiont genome. *Proceedings of the National Academy of Sciences* **113**, 15036-15041 (2016).
128 Stouthamer, R., Russell, J. E., Vavre, F. & Nunney, L. Intragenomic conflict in populations infected by parthenogenesis inducing Wolbachia ends with irreversible loss of sexual reproduction. *BMC evolutionary biology* **10**, 229 (2010).
129 Becking, T. *et al.* Sex chromosomes control vertical transmission of feminizing Wolbachia symbionts in an isopod. *bioRxiv*, 747444, doi:10.1101/747444 (2019).
130 Kageyama, D., Narita, S. & Watanabe, M. Insect sex determination manipulated by their endosymbionts: incidences, mechanisms and implications. *Insects* **3**, 161-199 (2012).
131 Morimoto, S., Nakai, M., Ono, A. & Kunimi, Y. Late male-killing phenomenon found in a Japanese population of the oriental tea tortrix, Homona magnanima (Lepidoptera: Tortricidae). *Heredity* **87**, 435 (2001).
132 Duron, O. *et al.* Origin, acquisition and diversification of heritable bacterial endosymbionts in louse flies and bat flies. *Molecular ecology* **23**, 2105-2117 (2014).
133 Hosokawa, T., Kikuchi, Y., Shimada, M. & Fukatsu, T. Obligate symbiont involved in pest status of host insect. *Proceedings of the Royal Society B: Biological Sciences* **274**, 1979-1984 (2007).
134 Oliver, K. M., Degnan, P. H., Burke, G. R. & Moran, N. A. Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annual review of entomology* **55**, 247-266 (2010).
135 Morse, S., Dick, C. W., Patterson, B. D. & Dittmar, K. Some like it hot–Evolution and ecology of novel endosymbionts in bat flies of cave-roosting bats (Hippoboscoidea, Nycterophiliinae). *Applied and environmental microbiology*, AEM. 02455-02412 (2012).
136 Brumin, M., Kontsedalov, S. & Ghanim, M. Rickettsia influences thermotolerance in the whitefly Bemisia tabaci B biotype. *Insect Science* **18**, 57-66 (2011).

137 Truitt, A. M., Kapun, M., Kaur, R. & Miller, W. J. Wolbachia modifies thermal preference in Drosophila melanogaster. *Environ Microbiol*, doi:10.1111/1462-2920.14347 (2018).
138 Corbin, C., Heyworth, E. R., Ferrari, J. & Hurst, G. D. Heritable symbionts in a world of varying temperature. *Heredity* **118**, 10 (2017).
139 Baião, G. C., Schneider, D. I., Miller, W. J. & Klasson, L. The effect of Wolbachia on gene expression in Drosophila paulistorum and its implications for symbiont-induced host speciation. *BMC Genomics* **20**, 465 (2019).
140 Zheng, Y. *et al.* Differentially expressed profiles in the larval testes of Wolbachia infected and uninfected Drosophila. *BMC genomics* **12**, 595 (2011).
141 Kupper, M., Stigloher, C., Feldhaar, H. & Gross, R. Distribution of the obligate endosymbiont Blochmannia floridanus and expression analysis of putative immune genes in ovaries of the carpenter ant Camponotus floridanus. *Arthropod structure & development* **45**, 475-487 (2016).
142 Negri, I. *et al.* Unravelling the Wolbachia evolutionary role: the reprogramming of the host genomic imprinting. *Proceedings of the Royal Society B: Biological Sciences* **276**, 2485-2491 (2009).
143 Wang, J. & Aksoy, S. PGRP-LB is a maternally transmitted immune milk protein that influences symbiosis and parasitism in tsetse's offspring. *Proceedings of the National Academy of Sciences* **109**, 10552-10557 (2012).
144 Papafotiou, G., Oehler, S., Savakis, C. & Bourtzis, K. Regulation of Wolbachia ankyrin domain encoding genes in Drosophila gonads. *Research in microbiology* **162**, 764-772 (2011).
145 Yixin, H. Y. *et al.* Infection with a virulent strain of Wolbachia disrupts genome wide-patterns of cytosine methylation in the mosquito Aedes aegypti. *PLoS One* **8**, e66482 (2013).
146 Bhattacharya, T., Newton, I. L. & Hardy, R. W. Wolbachia elevates host methyltransferase expression to block an RNA virus early during infection. *PLoS pathogens* **13**, e1006427 (2017).
147 Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. & Ishikawa, H. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. *Nature* **407**, 81 (2000).
148 Husnik, F. & McCutcheon, J. P. Functional horizontal gene transfer from bacteria to eukaryotes. *Nature Reviews Microbiology* **16**, 67 (2018).
149 Moran, N. A. & Jarvik, T. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science (New York, N.Y.)* **328**, 624-627 (2010).
150 Verster, K. I. *et al.* Horizontal transfer of bacterial cytolethal distending toxin B genes to insects. *Molecular Biology and Evolution* (2019).
151 Metcalf, J. A., Funkhouser-Jones, L. J., Brileya, K., Reysenbach, A.-L. & Bordenstein, S. R. Antibacterial gene transfer across the tree of life. *Elife* **3**, e04266 (2014).
152 Funkhouser-Jones, L. J. *et al.* Wolbachia co-infection in a hybrid zone: discovery of horizontal gene transfers from two Wolbachia supergroups into an animal genome. *PeerJ* **3**, e1479 (2015).
153 Hotopp, J. C. D. *et al.* Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science (New York, N.Y.)* **317**, 1753-1756 (2007).
154 Wybouw, N. *et al.* A gene horizontally transferred from bacteria protects arthropods from host plant cyanide poisoning. *Elife* **3**, e02365 (2014).

155 Blondel, L., Jones, T. E. & Extavour, C. G. Bacterial contribution to genesis of the novel germ line determinant oskar. *bioRxiv*, 453514 (2018).
156 Selman, M. *et al.* Acquisition of an animal gene by microsporidian intracellular parasites. *Current Biology* **21**, R576-R577 (2011).
157 Hughes, A. L. & Friedman, R. Genome-wide survey for genes horizontally transferred from cellular organisms to baculoviruses. *Molecular biology and evolution* **20**, 979-987 (2003).
158 LePage, D. P. *et al.* Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility. *Nature* **543**, 243-247, doi:10.1038/nature21391 (2017).
159 Beckmann, J. F., Ronau, J. A. & Hochstrasser, M. A Wolbachia deubiquitylating enzyme induces cytoplasmic incompatibility. *Nature microbiology* **2**, 17007 (2017).
160 Perlmutter, J. I. *et al.* The phage gene wmk is a candidate for male killing by a bacterial endosymbiont. *PLOS Pathogens* (2019).
161 Shropshire, J. D., On, J., Layton, E. M., Zhou, H. & Bordenstein, S. R. One prophage WO gene rescues cytoplasmic incompatibility in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 4987-4991, doi:10.1073/pnas.1800650115 (2018).
162 Hurst, G. D. & Jiggins, F. M. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B: Biological Sciences* **272**, 1525-1534 (2005).
163 Jiggins, F. M. & Tinsley, M. C. An ancient mitochondrial polymorphism in Adalia bipunctata linked to a sex-ratio-distorting bacterium. *Genetics* **171**, 1115-1124 (2005).
164 Ironside, J. E., Dunn, A., Rollinson, D. & Smith, J. Association with host mitochondrial haplotypes suggests that feminizing microsporidia lack horizontal transmission. *Journal of evolutionary biology* **16**, 1077-1083 (2003).
165 Charlat, S. *et al.* The joint evolutionary histories of Wolbachia and mitochondria in Hypolimnas bolina. *BMC evolutionary biology* **9**, 64 (2009).
166 Cariou, M., Duret, L. & Charlat, S. The global impact of Wolbachia on mitochondrial diversity and evolution. *Journal of evolutionary biology* **30**, 2204-2210 (2017).
167 Jiggins, F. M. Male-killing Wolbachia and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. *Genetics* **164**, 5-12 (2003).
168 Kapantaidaki, D. E. *et al.* Low levels of mitochondrial DNA and symbiont diversity in the worldwide agricultural pest, the greenhouse whitefly Trialeurodes vaporariorum (Hemiptera: Aleyrodidae). *Journal of Heredity* **106**, 80-92 (2014).
169 Beninati, T. *et al.* Absence of the symbiont Candidatus Midichloria mitochondrii in the mitochondria of the tick Ixodes holocyclus. *FEMS microbiology letters* **299**, 241-247 (2009).
170 Van Ham, R. C. *et al.* Reductive genome evolution in Buchnera aphidicola. *Proceedings of the National Academy of Sciences* **100**, 581-586 (2003).
171 Funk, D. J., Wernegreen, J. J. & Moran, N. A. Intraspecific variation in symbiont genomes: bottlenecks and the aphid-Buchnera association. *Genetics* **157**, 477-489 (2001).
172 Abbot, P. & Moran, N. A. Extremely low levels of genetic polymorphism in endosymbionts (Buchnera) of aphids (Pemphigus). *Mol Ecol* **11**, 2649-2660 (2002).
173 Moran, N. A. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences* **93**, 2873-2878 (1996).

174 Charlat, S. *et al.* Male-killing bacteria trigger a cycle of increasing male fatigue and female promiscuity. *Current Biology* **17**, 273-277 (2007).
175 Jaenike, J., Dyer, K. A., Cornish, C. & Minhas, M. S. Asymmetrical reinforcement and Wolbachia infection in Drosophila. *PLoS Biol* **4**, e325, doi:10.1371/journal.pbio.0040325 (2006).
176 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).
177 Burand, J. P., Tan, W., Kim, W., Nojima, S. & Roelofs, W. Infection with the insect virus Hz-2v alters mating behavior and pheromone production in female Helicoverpa zea moths. *Journal of insect science (Online)* **5**, 6, doi:10.1093/jis/5.1.6 (2005).
178 Brucker, R. M. & Bordenstein, S. R. Speciation by symbiosis. *Trends Ecol Evol* **27**, 443-451, doi:10.1016/j.tree.2012.03.011 (2012).
179 Miller, W. J., Ehrman, L. & Schneider, D. Infectious speciation revisited: impact of symbiont-depletion on female fitness and mating behavior of Drosophila paulistorum. *PLoS Pathogens* **6**, e1001214 (2010).
180 Krueger, C. M., Degrugillier, M. E. & Narang, S. K. Size Difference among 16S rRNA Genes from Endosymbiontic Bacteria Found in Testes of Heliothis virescens, H. subflexa,(Lepidoptera: Noctuidae) and Backcross Sterile Male Moths. *Florida Entomologist*, 382-390 (1993).
181 Bordenstein, S. R., O'Hara, F. P. & Werren, J. H. Wolbachia-induced incompatibility precedes other hybrid incompatibilities in Nasonia. *Nature* **409**, 707-710, doi:10.1038/35055543 (2001).
182 Gebiola, M., Kelly, S. E., Hammerstein, P., Giorgini, M. & Hunter, M. S. "Darwin's corollary" and cytoplasmic incompatibility induced by Cardinium may contribute to speciation in Encarsia wasps (Hymenoptera: Aphelinidae). *Evolution* **70**, 2447-2458, doi:10.1111/evo.13037 (2016).
183 O'Neill, S. L. *et al.* Scaled deployment of Wolbachia to protect the community from dengue and other Aedes transmitted arboviruses. *Gates open research* **2** (2018).
184 Fierer, N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology* **15**, 579 (2017).
185 Byrd, A. L., Belkaid, Y. & Segre, J. A. The human skin microbiome. *Nature Reviews Microbiology* **16**, 143 (2018).
186 van Oppen, M. J. & Blackall, L. L. Coral microbiome dynamics, functions and design in a changing world. *Nature Reviews Microbiology*, 1 (2019).
187 Gibson, J. *et al.* Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences* **111**, 8007-8012 (2014).
188 Degli Esposti, M. & Romero, E. M. The functional microbiome of arthropods. *PLoS One* **12**, e0176573 (2017).
189 Weinert, L. A., Araujo-Jnr, E. V., Ahmed, M. Z. & Welch, J. J. The incidence of bacterial endosymbionts in terrestrial arthropods. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20150249 (2015).
190 Mann, R. S., Pelz-Stelinski, K., Hermann, S. L., Tiwari, S. & Stelinski, L. L. Sexual transmission of a plant pathogenic bacterium, Candidatus Liberibacter asiaticus, between

conspecific insect vectors during mating. *PLoS One* **6**, e29197, doi:10.1371/journal.pone.0029197 (2011).
191 Ravel, J. *et al.* Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences of the United States of America* **108 Suppl 1**, 4680-4687, doi:10.1073/pnas.1002611107 (2011).
192 Bradford, L. L. & Ravel, J. The vaginal mycobiome: A contemporary perspective on fungi in women's health and diseases. *Virulence* **8**, 342-351 (2017).
193 Wylie, K. M. *et al.* The vaginal eukaryotic DNA virome and preterm birth. *American journal of obstetrics and gynecology* **219**, 189. e181-189. e112 (2018).
194 Aagaard, K. *et al.* A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy. *PLoS One* **7**, e36466, doi:10.1371/journal.pone.0036466 (2012).
195 Zhang, R. *et al.* Qualitative and semiquantitative analysis of Lactobacillus species in the vaginas of healthy fertile and postmenopausal Chinese women. *Journal of medical microbiology* **61**, 729-739, doi:10.1099/jmm.0.038687-0 (2012).
196 Shiraishi, T. *et al.* Influence of menstruation on the microbiota of healthy women's labia minora as analyzed using a 16S rRNA gene-based clone library method. *Japanese journal of infectious diseases* **64**, 76-80 (2011).
197 Noyes, N., Cho, K.-C., Ravel, J., Forney, L. J. & Abdo, Z. Associations between sexual habits, menstrual hygiene practices, demographics and the vaginal microbiome as revealed by Bayesian network analysis. *PloS one* **13**, e0191625 (2018).
198 Hunt, K. M. *et al.* Characterization of the diversity and temporal stability of bacterial communities in human milk. *PLoS One* **6**, e21313, doi:10.1371/journal.pone.0021313 (2011).
199 Hochreiter, W. W., Duncan, J. L. & Schaeffer, A. J. Evaluation of the bacterial flora of the prostate using a 16S rRNA gene based polymerase chain reaction. *The Journal of urology* **163**, 127-130 (2000).
200 Porter, C. M., Shrestha, E., Peiffer, L. B. & Sfanos, K. S. The microbiome in prostate inflammation and prostate cancer. *Prostate cancer and prostatic diseases* **21**, 345-354, doi:10.1038/s41391-018-0041-1 (2018).
201 Liu, C. M. *et al.* Male circumcision significantly reduces prevalence and load of genital anaerobic bacteria. *mBio* **4**, e00076-00013 (2013).
202 Nelson, D. E. *et al.* Bacterial communities of the coronal sulcus and distal urethra of adolescent males. *PloS one* **7**, e36298 (2012).
203 Mändar, R. *et al.* Seminal microbiome in men with and without prostatitis. *International Journal of Urology* **24**, 211-216 (2017).
204 Pellati, D. *et al.* Genital tract infections and infertility. *Eur J Obstet Gynecol Reprod Biol* **140**, 3-11, doi:10.1016/j.ejogrb.2008.03.009 (2008).
205 Payne, M. S. & Bayatibojakhi, S. Exploring preterm birth as a polymicrobial disease: an overview of the uterine microbiome. *Front Immunol* **5**, 595, doi:10.3389/fimmu.2014.00595 (2014).
206 Onderdonk, A. B., Delaney, M. L. & Fichorova, R. N. The Human Microbiome during Bacterial Vaginosis. *Clinical microbiology reviews* **29**, 223-238, doi:10.1128/cmr.00075-15 (2016).
207 Weng, S. L. *et al.* Bacterial communities in semen from men of infertile couples: metagenomic sequencing reveals relationships of seminal microbiota to semen quality. *PLoS One* **9**, e110152, doi:10.1371/journal.pone.0110152 (2014).

208 Sastry, K. S. *Seed-borne plant virus diseases*. (Springer Science & Business Media, 2013).
209 Saikkonen, K., Faeth, S. H., Helander, M. & Sullivan, T. Fungal endophytes: a continuum of interactions with host plants. *Annual review of Ecology and Systematics* **29**, 319-343 (1998).
210 Gao, F.-k., Dai, C.-c. & Liu, X.-z. Mechanisms of fungal endophytes in plant protection against pathogens. *African Journal of Microbiology Research* **4**, 1346-1351 (2010).
211 Truyens, S., Weyens, N., Cuypers, A. & Vangronsveld, J. Bacterial seed endophytes: genera, vertical transmission and interaction with plants. *Environ Microbiol* **7**, 40-50, doi:10.1111/1758-2229.12181 (2014).
212 Cankar, K., Kraigher, H., Ravnikar, M. & Rupnik, M. Bacterial endophytes from seeds of Norway spruce (Picea abies L. Karst). *FEMS Microbiol Lett* **244**, 341-345, doi:10.1016/j.femsle.2005.02.008 (2005).
213 Compant, S., Mitter, B., Colli-Mull, J. G., Gangl, H. & Sessitsch, A. Endophytes of grapevine flowers, berries, and seeds: identification of cultivable bacteria, comparison with other plant parts, and visualization of niches of colonization. *Microb Ecol* **62**, 188-197, doi:10.1007/s00248-011-9883-y (2011).
214 Kaga, H. *et al.* Rice seeds as sources of endophytic bacteria. *Microbes and environments* **24**, 154-162 (2009).
215 Lopez-Lopez, A., Rogel, M. A., Ormeno-Orrillo, E., Martinez-Romero, J. & Martinez-Romero, E. Phaseolus vulgaris seed-borne endophytic community with novel bacterial species such as Rhizobium endophyticum sp. nov. *Syst Appl Microbiol* **33**, 322-327, doi:10.1016/j.syapm.2010.07.005 (2010).
216 Mundt, J. O. & Hinkle, N. F. Bacteria within ovules and seeds. *Applied and environmental microbiology* **32**, 694-698 (1976).
217 Mano, H. *et al.* Culturable surface and endophytic bacterial flora of the maturing seeds of rice plants (Oryza sativa) cultivated in a paddy field. *Microbes and Environments* **21**, 86-100 (2006).
218 Shade, A., McManus, P. S. & Handelsman, J. Unexpected diversity during community succession in the apple flower microbiome. *mBio* **4**, e00602-00612 (2013).
219 Obersteiner, A. *et al.* Pollen-associated microbiome correlates with pollution parameters and the allergenicity of pollen. *PLoS One* **11**, e0149545 (2016).
220 Ambika Manirajan, B. *et al.* Bacterial microbiota associated with flower pollen is influenced by pollination type, and shows a high degree of diversity and species-specificity. *Environmental microbiology* **18**, 5161-5174 (2016).
221 Manirajan, B. A. *et al.* Diversity, specificity, co-occurrence and hub taxa of the bacterial–fungal pollen microbiome. *FEMS microbiology ecology* (2018).
222 Sugio, A. *et al.* Diverse targets of phytoplasma effectors: from plant development to defense against insects. *Annu Rev Phytopathol* **49**, 175-195, doi:10.1146/annurev-phyto-072910-095323 (2011).
223 MacLean, A. M. *et al.* Phytoplasma effector SAP54 hijacks plant reproduction by degrading MADS-box proteins and promotes insect colonization in a RAD23-dependent manner. *PLoS Biol* **12**, e1001835, doi:10.1371/journal.pbio.1001835 (2014).
224 Sugio, A., Kingdom, H. N., MacLean, A. M., Grieve, V. M. & Hogenhout, S. A. Phytoplasma protein effector SAP11 enhances insect vector reproduction by manipulating plant development and defense hormone biosynthesis. *Proceedings of the National*

Academy of Sciences of the United States of America **108**, E1254-1263, doi:10.1073/pnas.1105664108 (2011).
225 Puente, M. E., Li, C. Y. & Bashan, Y. Endophytic bacteria in cacti seeds can improve the development of cactus seedlings. *Environ Exp Bot* **66**, 402-408, doi:10.1016/j.envexpbot.2009.04.007 (2009).
226 Puente, M. E., Li, C. Y. & Bashan, Y. Rock-degrading endophytic bacteria in cacti. *Environ Exp Bot* **66**, 389-401, doi:doi:10.1016/j.envexpbot.2009.04.010 (2009).
227 Sánchez-López, A. S. *et al.* Community structure and diversity of endophytic bacteria in seeds of three consecutive generations of Crotalaria pumila growing on metal mine residues. *Plant and Soil* **422**, 51-66 (2018).
228 Carlier, A. L. & Eberl, L. The eroded genome of a Psychotria leaf symbiont: hypotheses about lifestyle and interactions with its plant host. *Environmental microbiology* **14**, 2757-2769 (2012).
229 Partida-Martinez, L. P. & Hertweck, C. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature* **437**, 884 (2005).
230 Pemberton, C. Obituary: Hubert W. Simmonds 1877-1966. (1966).
231 Simmonds, H. W. *My weapons had wings*. (Published privately by HW Simmonds under the patronage of the Fiji Society, 1964).
232 Simmonds, H. in *Proc. Entomol. Soc.*
233 Simmonds, H. in *Proc R Ent Soc Lond.* 29-32.
234 Simmonds, H. Further notes on Hypolimnas bolina L. *Fiji. Proc R Entomol Soc Lond* **5**, 75-77 (1930).
235 Hertig, M. & Wolbach, S. B. Studies on Rickettsia-Like Micro-Organisms in Insects. *The Journal of medical research* **44**, 329-374.327 (1924).
236 Poulton, E. B. WA Lamborn's Breeding Experiments upon Acræa encedon (Linn.), in the Lagos District of West Africa, 1910–1912. *Zoological Journal of the Linnean Society* **32**, 391-416 (1914).
237 Lusis, J. in *Dokl. Akad. Nauk SSSR.* 951-954.
238 Magni, G. 'Sex-ratio': a non-Mendelian character in Drosophila bifasciata. *Nature* **172**, 81-81 (1953).
239 Cavalcanti, A. & Falcao, D. (Caryologia, 1954).
240 Carson, H. A female producing strain of D. borealis Patterson. *Drosophila Information Service* **30**, 109-110 (1956).
241 Malogolowkin, C. & Poulson, D. Infective transfer of maternally inherited abnormal sex-ratio in Drosophila willistoni. *Science (New York, N.Y.)* **126**, 32-32 (1957).
242 Poulson, D. & Sakaguchi, B. Nature of" sex-ratio" agent in Drosophila. *Science (New York, N.Y.)* **133**, 1489-1490 (1961).
243 Sakaguchi, B. & Poulson, D. Interspecific transfer of the" sex-ratio" condition from Drosophila willistoni to D. melanogaster. *Genetics* **48**, 841 (1963).
244 Oishi, K. Spirochaete-mediated abnormal sex-ratio (SR) condition in Drosophila: a second virus associated with spirochaetes and its use in the study of the SR condition. *Genetics Research* **18**, 45-56 (1971).
245 Hurst, G. D. D. *et al.* 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).

246 Williamson, D. L. *et al.* Spiroplasma poulsonii sp. nov., a new species associated with male-lethality in Drosophila willistoni, a neotropical species of fruit fly. *International Journal of Systematic and Evolutionary Microbiology* **49**, 611-618 (1999).
247 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).
248 Harumoto, T. & Lemaitre, B. Male-killing toxin in a bacterial symbiont of Drosophila. *Nature* **557**, 252 (2018).
249 Perlmutter, J. I. & Bordenstein, S. R. Microbial Misandry: Discovery of a Spiroplasma Male-Killing Toxin. *Cell host & microbe* **23**, 689-690 (2018).
250 Kellen, W. R., Chapman, H. C., Clark, T. B. & Lindegren, J. E. Host-parasite relationships of some Thelohania from mosquitoes (Nosematidae: Microsporidia). *Journal of Invertebrate Pathology* **7**, 161-166 (1965).
251 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).
252 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).
253 Charlat, S. *et al.* 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).
254 Hayashi, M., Watanabe, M., Yukuhiro, F., Nomura, M. & Kageyama, D. A nightmare for males? A maternally transmitted male-killing bacterium and strong female bias in a green lacewing population. *PLoS One* **11**, e0155794 (2016).
255 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).
256 Darby, A. C. *et al.* Characteristics of the genome of Arsenophonus nasoniae, son-killer bacterium of the wasp Nasonia. *Insect Mol Biol* **19 Suppl 1**, 75-89, doi:10.1111/j.1365-2583.2009.00950.x (2010).
257 Roberts, L., Rapmund, G. & Gadigan, F. G. Sex ratios in Rickettsia tsutsugamushi-infected and noninfected colonies of Leptotrombidium (Acari: Trombiculidae). *Journal of medical entomology* **14**, 89-92 (1977).
258 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).
259 Hurst, G. D. & Majerus, M. E. Why do maternally inherited microorganisms kill males? *Heredity* **71**, 81-95, doi:10.1038/hdy.1993.110

(1993).
260 Lassy, C. W. & Karr, T. L. Cytological analysis of fertilization and early embryonic development in incompatible crosses of Drosophila simulans. *Mechanisms of development* **57**, 47-58, doi:10.1016/0925-4773(96)00527-8 (1996).
261 Majerus, M. E. The impact of male-killing bacteria on the evolution of aphidophagous coccinellids. *European Journal of Entomology* **103**, 1 (2006).
262 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).

263 Sheeley, S. & McAllister, B. Mobile male-killer: similar Wolbachia strains kill males of divergent Drosophila hosts. *Heredity* **102**, 286 (2009).
264 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).
265 Richardson, K. M., Schiffer, M., Griffin, P. C., Lee, S. F. & Hoffmann, A. A. Tropical Drosophila pandora carry Wolbachia infections causing cytoplasmic incompatibility or male killing. *Evolution* **70**, 1791-1802, doi:10.1111/evo.12981 (2016).
266 Jaenike, J. Spontaneous emergence of a new Wolbachia phenotype. *Evolution* **61**, 2244-2252, doi:10.1111/j.1558-5646.2007.00180.x (2007).
267 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).
268 Arai, H., Lin, S. R., Nakai, M., Kunimi, Y. & Inoue, M. N. Closely Related Male-Killing and Nonmale-Killing Wolbachia Strains in the Oriental Tea Tortrix Homona magnanima. *Microbial Ecology*, 1-10 (2019).
269 Sakamoto, H., Kageyama, D., Hoshizaki, S. & Ishikawa, Y. Sex-specific death in the Asian corn borer moth (Ostrinia furnacalis) infected with Wolbachia occurs across larval development. *Genome* **50**, 645-652 (2007).
270 Kageyama, D. *et al.* Wolbachia infection and an all-female trait in Ostrinia orientalis and Ostrinia zaguliaevi. *Entomologia experimentalis et applicata* **111**, 79-83 (2004).
271 Kageyama, D., Nishimura, G., Hoshizaki, S. & Ishikawa, Y. Two kinds of sex ratio distorters in a moth, Ostrinia scapulalis. *Genome* **46**, 974-982 (2003).
272 Graham, R. I. & Wilson, K. Male-killing Wolbachia and mitochondrial selective sweep in a migratory African insect. *BMC evolutionary biology* **12**, 204 (2012).
273 Majerus, M. E., Hinrich, J., Schulenburg, G. & Zakharov, I. A. Multiple causes of male-killing in a single sample of the two-spot ladybird, Adalia bipunctata (Coleoptera: Coccinellidae) from Moscow. *Heredity* **84**, 605 (2000).
274 von der Schulenburg, J. H. G. *et al.* Incidence of Male-Killing Rickettsiaspp.(α-Proteobacteria) in the Ten-Spot Ladybird BeetleAdalia decempunctata L.(Coleoptera: Coccinellidae). *Appl. Environ. Microbiol.* **67**, 270-277 (2001).
275 Lawson, E. T., Mousseau, T. A., Klaper, R., Hunter, M. D. & Werren, J. H. Rickettsia associated with male-killing in a buprestid beetle. *Heredity* **86**, 497 (2001).
276 Majerus, T. M. & Majerus, M. E. Discovery and identification of a male-killing agent in the Japanese ladybird Propylea japonica (Coleoptera: Coccinellidae). *BMC evolutionary biology* **10**, 37 (2010).
277 Hurst, G. D. *et al.* Invasion of one insect species, Adalia bipunctata, by two different male-killing bacteria. *Insect Mol Biol* **8**, 133-139 (1999).
278 Fukatsu, T., Tsuchida, T., Nikoh, N. & Koga, R. Spiroplasma symbiont of the pea aphid, Acyrthosiphon pisum (Insecta: Homoptera). *Applied and environmental microbiology* **67**, 1284-1291 (2001).
279 Tinsley, M. & Majerus, M. A new male-killing parasitism: Spiroplasma bacteria infect the ladybird beetle Anisosticta novemdecimpunctata (Coleoptera: Coccinellidae). *Parasitology* **132**, 757-765 (2006).

280 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).
281 Jiggins, F., Hurst, G., Jiggins, C., vd Schulenburg, J. & Majerus, M. The butterfly Danaus chrysippus is infected by a male-killing Spiroplasma bacterium. *Parasitology* **120**, 439-446 (2000).
282 MATSUKA, M., HASHI, H. & OKADA, I. Abnormal sex-ratio found in the lady beetle, Harmonia axyridis Pallas (Coleoptera: Coccinellidae). *Applied Entomology and Zoology* **10**, 84-89 (1975).
283 Tsugeno, Y. *et al.* Identification of an early male-killing agent in the oriental tea tortrix, Homona magnanima. *Journal of Heredity* **108**, 553-560 (2017).
284 Sanada-Morimura, S., Matsumura, M. & Noda, H. Male killing caused by a Spiroplasma symbiont in the small brown planthopper, Laodelphax striatellus. *Journal of Heredity* **104**, 821-829 (2013).
285 Niijima, K. & Nakajima, K. Abnormal sex-ratio in Menochilus sexmaculatus (Fabricius). *Bulletin of the Faculty of Agriculture-Tamagawa University (Japan)* (1981).
286 Tabata, J. *et al.* Male killing and incomplete inheritance of a novel Spiroplasma in the moth Ostrinia zaguliaevi. *Microbial ecology* **61**, 254-263 (2011).
287 Bentley, J. K., Veneti, Z., Heraty, J. & Hurst, G. D. The pathology of embryo death caused by the male-killing Spiroplasma bacterium in Drosophila nebulosa. *Bmc Biology* **5**, 9 (2007).
288 Montenegro, H., Hatadani, L., Medeiros, H. & Klaczko, L. Male killing in three species of the tripunctata radiation of Drosophila (Diptera: Drosophilidae). *Journal of Zoological Systematics and Evolutionary Research* **44**, 130-135 (2006).
289 Montenegro, H., Solferini, V. N., Klaczko, L. B. & Hurst, G. D. Male-killing Spiroplasma naturally infecting Drosophila melanogaster. *Insect Mol Biol* **14**, 281-287, doi:10.1111/j.1365-2583.2005.00558.x (2005).
290 Malogolowkin, C. Temperature effects on maternally inherited" sex-ratio" conditions in Drosophila willistoni and Drosophila equinoxialis. *The American Naturalist* **93**, 365-368 (1959).
291 Malogolowkin, C. Maternally inherited" sex-ratio" conditions in Drosophila willistoni and Drosophila paulistorum. *Genetics* **43**, 274 (1958).
292 Falcao, D. N. & Castro, L. E. "Sex-Ratio" in Drosophila prosaltans- a character due to interaction between nuclear genes and cytoplasmic factors. *The American Naturalist* **91**, 327-329 (1957).
293 Poulson, D. Further cases of maternal SR in Drosophila species. *Dros. Inf. Serv* **41**, 77 (1966).
294 Hurst, G. *et al.* Adonia variegata (Coleoptera: Coccinellidae) bears maternally inherited Flavobacteria that kill males only. *Parasitology* **118**, 125-134 (1999).
295 Majerus, M. E. & Majerus, T. M. Female-biased sex ratio due to male-killing in the Japanese ladybird Coccinula sinensis. *Ecological Entomology* **25**, 234-238 (2000).
296 Hurst, G. D. *et al.* The diversity of inherited parasites of insects: the male-killing agent of the ladybird beetle Coleomegilla maculata is a member of the Flavobacteria. *Genetics Research* **70**, 1-6 (1997).

297 Takahashi, F. & Kuwahara, Y. Studies on sex pheromones of Pyralididae. III. The inheritance of the abnormal sex ratio condition in a strain of the almond moth, Cadra cautella Walker (Phycitinae). *Botyu-Kagaku= Scientific Insect Control* **35** (1970).
298 Majerus, M. E. *Sex wars: genes, bacteria, and biased sex ratios*. (Princeton University Press, 2003).
299 CHANG, K. S., SHIRAISHI, T., NAKASUJI, F. & MORIMOTO, N. Abnormal sex-ratio condition in the walnut leaf beetle, Gastrolina depressa (Coleoptera: Chrysomelidae). *Applied Entomology and Zoology* **26**, 299-306 (1991).
300 Higashiura, Y., Ishihara, M. & Schaefer, P. W. Sex ratio distortion and severe inbreeding depression in the gypsy moth Lymantria dispar L. in Hokkaido, Japan. *Heredity* **83 (Pt 3)**, 290-297 (1999).
301 Groeters, F. R. Maternally inherited sex ratio distortion as a result of a male-killing agent in Spilostethus hospes (Hemiptera: Lygaeidae). *Heredity* **77**, 201 (1996).
302 Andreadis, T. G. Life cycle, epizootiology, and horizontal transmission of Amblyospora (Microspora: Amblyosporidae) in a univoltine mosquito, Aedes stimulans. *Journal of Invertebrate Pathology* **46**, 31-46 (1985).
303 Andreadis, T. G. & Hall, D. W. Significance of transovarial infections of Amblyospora sp.(Microspora: Thelohaniidae) in relation to parasite maintenance in the mosquito Culex salinarius. *Journal of Invertebrate Pathology* **34**, 152-157 (1979).
304 Hazard, E. I. & Weiser, J. Spores of Thelohania in adult female Anopheles: development and transovarial transmission, and redescriptions of T. legeri Hesse and T. obesa Kudo. *The Journal of protozoology* **15**, 817-823 (1968).
305 Avery, S. W. Horizontal transmission of Parathelohania obesa (Protozoa: microsporida) to Anopheles quadrimaculatus (Diptera: Culicidae). *Journal of Invertebrate Pathology* **53**, 424-426 (1989).
306 Kellen, W. & Wills, W. The transovarian transmission of Thelohania californica Kellen and Lipa in Culex tarsalis Coqulllett. *Journal of Insect Pathology* **4** (1962).
307 Avery, S. W. & Undeen, A. H. Horizontal transmission of Parathelohania anophelis to the copepod, Microcyclops varicans, and the mosquito, Anopheles quadrimaculatus. *Journal of invertebrate pathology* **56**, 98-105 (1990).
308 Nakamoto, N. *Analysis of the damage by dermestid beetles (Coleoptera: Dermestidae) in the sericultural industry and its control*, Ph. D. Thesis, Kyushu Univ.(in Japanese with English summary), (1984).
309 Robertson, J. Effect of supernumerary chromosomes on sex ratio in Calligrapha philadelphica L.(Coleoptera: Chrysomelidae). *Nature* **204**, 605-605 (1964).
310 Jackson, D. J. Observations on the biology of Caraphractus cinctus Walker (Hymenoptera: Mymaridae), a parasitoid of the eggs of Dytiscidae (Coleoptera) III. The adult life and sex ratio. *Transactions of the Royal Entomological Society of London* **118**, 23-49 (1966).
311 Lanier, G. & Wood, D. Controlled mating, karyology, morphology, and sex-ratio in the Dendroctonus ponderosae complex. *Annals of the Entomological Society of America* **61**, 517-526 (1968).
312 Geier, P. & Briese, D. The light-brown apple moth, Epiphyas postvittana (Walker): 4. Studies on population dynamics and injuriousness to apples in the Australian Capital Territory. *Australian journal of ecology* **5**, 63-93 (1980).
313 Earle, N. & Macfarlane, J. A unisexual strain of the salt-marsh caterpillar, Estigmene acrea. *Annals of the Entomological Society of America* **61**, 949-953 (1968).

314 Shull, A. F. An all-female strain of lady beetles with reversions to normal sex ratio. *The American Naturalist* **82**, 241-251 (1948).
315 Lanier, G. & Oliver, J. " Sex-Ratio" Condition: Unusual Mechanisms in Bark Beetles. *Science (New York, N.Y.)* **153**, 208-209 (1966).
316 Munro, J. in *Proceedings of the Entomological Society of London.* 33-34.
317 Leslie, J. F. A "sex-ratio" condition in Oncopeltus fasciatus. *Journal of Heredity* **75**, 260-264 (1984).
318 McCoy, R., Tsai, J. & Thomas, D. Occurrence of a spiroplasma in natural populations of the sharpshooter Oncometopia nigricans. *Phytopathology News* **12**, 217 (1978).
319 Ujiye, T. Abnormal sex-ratio in the population of Phyllonorycter sorbicola Kumata (Lepidoptera: Gracillariidae). *Jap. J. Appl. Entomol. Zool.* **25**, 244-248 (1981).
320 Bowden, S. Sex ratio in Pieris hybrids. *J. Lepid. Soc* (1966).
321 FEDERLEY, H. Sex-limited hereditary cancer in Lepidopterous larvae. *Hereditas* **22**, 193-216 (1936).
322 Brimacombe, L. C. All-female broods in field and laboratory populations of the Egyptian cotton leafworm, Spodoptera littoralis (Boisduval)(Lepidoptera: Noctuidae). *Bulletin of Entomological Research* **70**, 475-481 (1980).
323 Mitchell, R. The sex ratio of the spider mite Tetranychus urticae. *Entomologia experimentalis et applicata* **15**, 299-304 (1972).
324 Chamberlin, W. The bark and timber beetles of North America. *Oregon State Coop Association, Corvallis, OR* (1939).
325 Ikeda, H. The cytoplasmically-inherited "sex-ratio" condition in natural and experimental populations of Drosophila bifasciata. *Genetics* **65**, 311-333 (1970).
326 Arai, H. *et al.* Multiple infection and reproductive manipulations of Wolbachia in Homona magnanima (Lepidoptera: Tortricidae). *Microbial ecology* **77**, 257-266 (2019).
327 Engelstadter, J. & Hurst, G. D. The impact of male-killing bacteria on host evolutionary processes. *Genetics* **175**, 245-254, doi:10.1534/genetics.106.060921 (2007).
328 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).
329 Hornett, E. A. *et al.* 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).
330 Jaenike, J. & Dyer, K. No resistance to male-killing Wolbachia after thousands of years of infection. *Journal of evolutionary biology* **21**, 1570-1577 (2008).
331 Reynolds, L. A., Hornett, E. A., Jiggins, C. D. & Hurst, G. D. Suppression of Wolbachia-mediated male-killing in the butterfly Hypolimnas bolina involves a single genomic region. *PeerJ* **7**, e7677 (2019).
332 Werren, J. H. The coevolution of autosomal and cytoplasmic sex ratio factors. *Journal of Theoretical Biology* **124**, 317-334 (1987).
333 Dannowski, J., Flor, M., Telschow, A. & Hammerstein, P. The effect of sibmating on the infection dynamics of male-killing bacteria. *Evolution* **63**, 2525-2534, doi:10.1111/j.1558-5646.2009.00749.x (2009).
334 Elnagdy, S., Majerus, M. & HANDLEY, L. J. L. The value of an egg: resource reallocation in ladybirds (Coleoptera: Coccinellidae) infected with male-killing bacteria. *Journal of evolutionary biology* **24**, 2164-2172 (2011).

335 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).
336 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).
337 Longdon, B., Fabian, D. K., Hurst, G. D. & Jiggins, F. M. Male-killing Wolbachia do not protect Drosophila bifasciata against viral infection. *BMC microbiology* **12**, S8 (2012).
338 Hurst, G. D., Jiggins, F. M. & Robinson, S. J. What causes inefficient transmission of male-killing Wolbachia in Drosophila? *Heredity* **87**, 220 (2001).
339 Montenegro, H. & Klaczko, L. Low temperature cure of a male killing agent in Drosophila melanogaster. *Journal of invertebrate pathology* **86**, 50-51 (2004).
340 Anbutsu, H., Goto, S. & Fukatsu, T. High and low temperatures differently affect infection density and vertical transmission of male-killing Spiroplasma symbionts in Drosophila hosts. *Appl. Environ. Microbiol.* **74**, 6053-6059 (2008).
341 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).
342 Veneti, Z. *et al.* 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).
343 Cavalcanti, A. L., Falcao, D. & Castro, L. " Sex-Ratio" in Drosophila prosaltans-A Character Due to Interaction between Nuclear Genes and Cytoplasmic Factors. *The American Naturalist* **91**, 327-329 (1957).
344 Sakamoto, H. *et al.* Transinfection reveals the crucial importance of Wolbachia genotypes in determining the type of reproductive alteration in the host. *Genetical research* **85**, 205-210, doi:10.1017/s0016672305007573 (2005).
345 Hornett, E. A. *et al.* Evolution of male-killer suppression in a natural population. *PLoS Biol* **4**, e283, doi:10.1371/journal.pbio.0040283 (2006).
346 Normark, B. B. Haplodiploidy as an outcome of coevolution between male-killing cytoplasmic elements and their hosts. *Evolution* **58**, 790-798 (2004).
347 Ferree, P. M., Avery, A., Azpurua, J., Wilkes, T. & Werren, J. H. A bacterium targets maternally inherited centrosomes to kill males in Nasonia. *Current biology : CB* **18**, 1409-1414, doi:10.1016/j.cub.2008.07.093 (2008).
348 Cordaux, R., Bouchon, D. & Grève, P. The impact of endosymbionts on the evolution of host sex-determination mechanisms. *Trends in Genetics* **27**, 332-341 (2011).
349 Werren, J. H. & Beukeboom, L. W. Sex determination, sex ratios, and genetic conflict. *Annual Review of Ecology and Systematics* **29**, 233-261 (1998).
350 Bell, L. R., Maine, E. M., Schedl, P. & Cline, T. W. Sex-lethal, a Drosophila sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell* **55**, 1037-1046 (1988).
351 Bopp, D., Calhoun, G., Horabin, J. I., Samuels, M. & Schedl, P. Sex-specific control of Sex-lethal is a conserved mechanism for sex determination in the genus Drosophila. *Development (Cambridge, England)* **122**, 971-982 (1996).

352 Ruiz, M. F., Esteban, M. R., Donoro, C., Goday, C. & Sánchez, L. Evolution of dosage compensation in Diptera: the gene maleless implements dosage compensation in Drosophila (Brachycera suborder) but its homolog in Sciara (Nematocera suborder) appears to play no role in dosage compensation. *Genetics* **156**, 1853-1865 (2000).
353 Verhulst, E. C., van de Zande, L. & Beukeboom, L. W. Insect sex determination: it all evolves around transformer. *Current opinion in genetics & development* **20**, 376-383 (2010).
354 Pastok, D., Atkinson, D. & Hurst, G. D. Assessing the impact of male-killing bacteria on the spread of a sexually transmitted infection. *Animal behaviour* **107**, 41-48 (2015).
355 Randerson, J. P., Jiggins, F. M. & Hurst, L. D. Male killing can select for male mate choice: a novel solution to the paradox of the lek. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **267**, 867-874 (2000).
356 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).
357 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).
358 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).
359 Veneti, Z., Bentley, J. K., Koana, T., Braig, H. R. & Hurst, G. D. A functional dosage compensation complex required for male killing in Drosophila. *Science (New York, N.Y.)* **307**, 1461-1463, doi:10.1126/science.1107182 (2005).
360 Harumoto, T., Anbutsu, H., Lemaitre, B. & Fukatsu, T. Male-killing symbiont damages host's dosage-compensated sex chromosome to induce embryonic apoptosis. *Nature communications* **7**, 12781, doi:10.1038/ncomms12781 (2016).
361 Harumoto, T., Anbutsu, H. & Fukatsu, T. Male-killing Spiroplasma induces sex-specific cell death via host apoptotic pathway. *PLoS pathogens* **10**, e1003956 (2014).
362 Kageyama, D., Ohno, S., Hoshizaki, S. & Ishikawa, Y. Sexual mosaics induced by tetracycline treatment in the Wolbachia-infected adzuki bean borer, Ostrinia scapulalis. *Genome* **46**, 983-989 (2003).
363 Kageyama, D. & Traut, W. Opposite sex–specific effects of Wolbachia and interference with the sex determination of its host Ostrinia scapulalis. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**, 251-258 (2004).
364 Sugimoto, T. N., Fujii, T., Kayukawa, T., Sakamoto, H. & Ishikawa, Y. Expression of a doublesex homologue is altered in sexual mosaics of Ostrinia scapulalis moths infected with Wolbachia. *Insect biochemistry and molecular biology* **40**, 847-854 (2010).
365 Sugimoto, T. N. & Ishikawa, Y. A male-killing Wolbachia carries a feminizing factor and is associated with degradation of the sex-determining system of its host. *Biol Lett* **8**, 412-415, doi:10.1098/rsbl.2011.1114 (2012).
366 Fukui, T. *et al.* The Endosymbiotic Bacterium Wolbachia Selectively Kills Male Hosts by Targeting the Masculinizing Gene. *PLoS Pathog* **11**, e1005048, doi:10.1371/journal.ppat.1005048 (2015).

367 Sugimoto, T. N., Kayukawa, T., Shinoda, T., Ishikawa, Y. & Tsuchida, T. Misdirection of dosage compensation underlies bidirectional sex-specific death in Wolbachia-infected Ostrinia scapulalis. *Insect Biochem Mol Biol* **66**, 72-76, doi:10.1016/j.ibmb.2015.10.001 (2015).
368 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).
369 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, 1 (2020).
370 LePage, D. & Bordenstein, S. R. Wolbachia: Can we save lives with a great pandemic? *Trends Parasitol* **29**, 385-393, doi:10.1016/j.pt.2013.06.003 (2013).
371 Mains, J. W., Kelly, P. H., Dobson, K. L., Petrie, W. D. & Dobson, S. L. Localized Control of Aedes aegypti (Diptera: Culicidae) in Miami, FL, via Inundative Releases of Wolbachia-Infected Male Mosquitoes. *Journal of medical entomology* (2019).
372 Ford, S. A. *et al.* Selection on Aedes aegypti alters Wolbachia-mediated dengue virus blocking and fitness. *Nature microbiology*, 1-8 (2019).
373 Berec, L., Maxin, D. & Bernhauerova, V. Male-killing bacteria as agents of insect pest control. *Journal of Applied Ecology* (2016).
374 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).
375 Shropshire, J. D. & Bordenstein, S. R. Speciation by Symbiosis: the Microbiome and Behavior. *mBio* **7**, e01785, doi:10.1128/mBio.01785-15 (2016).
376 O'Connor, L. *et al.* Open release of male mosquitoes infected with a wolbachia biopesticide: field performance and infection containment. *PLoS Negl Trop Dis* **6**, e1797, doi:10.1371/journal.pntd.0001797 (2012).
377 Walker, T. *et al.* The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations. *Nature* **476**, 450-453, doi:10.1038/nature10355 (2011).
378 Zabalou, S. *et al.* Wolbachia-induced cytoplasmic incompatibility as a means for insect pest population control. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15042-15045, doi:10.1073/pnas.0403853101 (2004).
379 Ishmael, N. *et al.* Extensive genomic diversity of closely related Wolbachia strains. *Microbiology* **155**, 2211 (2009).
380 Beckmann, J. F. & Fallon, A. M. Detection of the Wolbachia protein WPIP0282 in mosquito spermathecae: implications for cytoplasmic incompatibility. *Insect Biochem Mol Biol* **43**, 867-878, doi:10.1016/j.ibmb.2013.07.002 (2013).
381 Bossan, B., Koehncke, A. & Hammerstein, P. A new model and method for understanding Wolbachia-induced cytoplasmic incompatibility. *PLoS One* **6**, e19757 (2011).
382 Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567-580, doi:10.1006/jmbi.2000.4315 (2001).
383 Lorenzen, M. D. *et al.* The maternal-effect, selfish genetic element Medea is associated with a composite Tc1 transposon. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 10085-10089, doi:10.1073/pnas.0800444105 (2008).
384 Zabalou, S. *et al.* Multiple rescue factors within a Wolbachia strain. *Genetics* **178**, 2145-2160, doi:10.1534/genetics.107.086488 (2008).

385 Poinsot, D., Bourtzis, K., Markakis, G., Savakis, C. & Mercot, H. Wolbachia transfer from Drosophila melanogaster into D. simulans: Host effect and cytoplasmic incompatibility relationships. *Genetics* **150**, 227-237 (1998).
386 Reynolds, K. T. & Hoffmann, A. A. Male age, host effects and the weak expression or non-expression of cytoplasmic incompatibility in Drosophila strains infected by maternally transmitted Wolbachia. *Genetics Research* **80**, 79-87 (2002).
387 Gutzwiller, F. *et al.* Dynamics of Wolbachia pipientis gene expression across the Drosophila melanogaster life cycle. *G3: Genes, Genomes, Genetics* **5**, 2843-2856 (2015).
388 Yamada, R., Floate, K. D., Riegler, M. & O'Neill, S. L. Male development time influences the strength of Wolbachia-induced cytoplasmic incompatibility expression in Drosophila melanogaster. *Genetics* **177**, 801-808, doi:10.1534/genetics.106.068486 (2007).
389 Rorth, P. Gal4 in the Drosophila female germline. *Mechanisms of development* **78**, 113-118, doi:10.1016/s0925-4773(98)00157-9 (1998).
390 White-Cooper, H. Tissue, cell type and stage-specific ectopic gene expression and RNAi induction in the Drosophila testis. *Spermatogenesis* **2**, 11-22, doi:10.4161/spmg.19088 (2012).
391 Serbus, L. R., Casper-Lindley, C., Landmann, F. & Sullivan, W. The genetics and cell biology of Wolbachia-host interactions. *Annual review of genetics* **42**, 683-707 (2008).
392 Landmann, F., Orsi, G. A., Loppin, B. & Sullivan, W. Wolbachia-mediated cytoplasmic incompatibility is associated with impaired histone deposition in the male pronucleus. *PLoS pathogens* **5**, e1000343 (2009).
393 Callaini, G., Riparbelli, M. G., Giordano, R. & Dallai, R. Mitotic Defects Associated with Cytoplasmic Incompatibility inDrosophila simulans. *Journal of Invertebrate Pathology* **67**, 55-64 (1996).
394 Wright, J. D. & Barr, A. R. Wolbachia and the normal and incompatible eggs of Aedes polynesiensis (Diptera: Culicidae). *Journal of Invertebrate Pathology* **38**, 409-418 (1981).
395 Duron, O. & Weill, M. Wolbachia infection influences the development of Culex pipiens embryo in incompatible crosses. *Heredity* **96**, 493-500, doi:10.1038/sj.hdy.6800831 (2006).
396 Bate, M., Martinez Arias, A. & Ingham, P. The Development of Drosophila melanogaster. *Trends in Genetics* **10**, 299-299 (1994).
397 Zug, R. & Hammerstein, P. Still a host of hosts for Wolbachia: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS One* **7**, e38544, doi:10.1371/journal.pone.0038544 (2012).
398 Vallenet, D. *et al.* MicroScope: a platform for microbial genome annotation and comparative genomics. *Database : the journal of biological databases and curation* **2009**, bap021, doi:10.1093/database/bap021 (2009).
399 Wu, M. *et al.* Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* **2**, E69, doi:10.1371/journal.pbio.0020069 (2004).
400 Klasson, L. *et al.* The mosaic genome structure of the Wolbachia wRi strain infecting Drosophila simulans. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 5725-5730, doi:10.1073/pnas.0810753106 (2009).
401 Klasson, L. *et al.* Genome evolution of Wolbachia strain wPip from the Culex pipiens group. *Mol Biol Evol* **25**, 1877-1887, doi:10.1093/molbev/msn133 (2008).

402 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).
403 Foster, J. *et al.* The Wolbachia genome of Brugia malayi: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol* **3**, e121, doi:10.1371/journal.pbio.0030121 (2005).
404 Werren, J. H. & Loehlin, D. W. Rearing Sarcophaga bullata fly hosts for Nasonia (parasitoid wasp). *Cold Spring Harbor protocols* **2009**, pdb.prot5308, doi:10.1101/pdb.prot5308 (2009).
405 Bordenstein, S. R. & Bordenstein, S. R. Temperature affects the tripartite interactions between bacteriophage WO, Wolbachia, and cytoplasmic incompatibility. *PLoS One* **6**, e29106, doi:10.1371/journal.pone.0029106 (2011).
406 Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)* **28**, 1647-1649, doi:10.1093/bioinformatics/bts199 (2012).
407 Beckmann, J. F., Markowski, T. W., Witthuhn, B. A. & Fallon, A. M. Detection of the Wolbachia-encoded DNA binding protein, HU beta, in mosquito gonads. *Insect Biochem Mol Biol* **43**, 272-279, doi:10.1016/j.ibmb.2012.12.007 (2013).
408 Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic acids research* **36**, W5-9, doi:10.1093/nar/gkn201 (2008).
409 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797 (2004).
410 Hurvich, C. M. & Tsai, C. L. A corrected Akaike information criterion for vector autoregressive model selection. *Journal of time series analysis* **14**, 271-279 (1993).
411 Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics (Oxford, England)* **21**, 2104-2105, doi:10.1093/bioinformatics/bti263 (2005).
412 Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology* **61**, 539-542, doi:10.1093/sysbio/sys029 (2012).
413 Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic acids research* **44**, D279-285, doi:10.1093/nar/gkv1344 (2016).
414 Letunic, I., Doerks, T. & Bork, P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic acids research* **40**, D302-305, doi:10.1093/nar/gkr931 (2012).
415 Casiraghi, M., Anderson, T. J., Bandi, C., Bazzocchi, C. & Genchi, C. A phylogenetic analysis of filarial nematodes: comparison with the phylogeny of Wolbachia endosymbionts. *Parasitology* **122 Pt 1**, 93-103, doi:10.1017/s0031182000007149 (2001).
416 Chatzispyrou, I. A., Held, N. M., Mouchiroud, L., Auwerx, J. & Houtkooper, R. H. Tetracycline antibiotics impair mitochondrial function and its experimental use confounds research. *Cancer research* **75**, 4446-4449, doi:10.1158/0008-5472.Can-15-1626 (2015).
417 Ferguson, S. B., Blundon, M. A., Klovstad, M. S. & Schupbach, T. Modulation of gurken translation by insulin and TOR signaling in Drosophila. *J Cell Sci* **125**, 1407-1419, doi:10.1242/jcs.090381 (2012).
418 Groth, A. C., Fish, M., Nusse, R. & Calos, M. P. Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics* **166**, 1775-1782 (2004).

419 Southall, T. D., Elliott, D. A. & Brand, A. H. The GAL4 System: A Versatile Toolkit for Gene Expression in Drosophila. *CSH Protoc* **2008**, pdb top49, doi:10.1101/pdb.top49 (2008).
420 LePage, D. P., Jernigan, K. K. & Bordenstein, S. R. The relative importance of DNA methylation and Dnmt2-mediated epigenetic regulation on Wolbachia densities and cytoplasmic incompatibility. *PeerJ* **2**, e678, doi:10.7717/peerj.678 (2014).
421 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9**, 671-675, doi:10.1038/nmeth.2089 (2012).
422 Vizcaino, J. A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic acids research* **44**, D447-456, doi:10.1093/nar/gkv1145 (2016).
423 Ferri, E. *et al.* New insights into the evolution of Wolbachia infections in filarial nematodes inferred from a large range of screened species. *PLoS One* **6**, e20843, doi:10.1371/journal.pone.0020843 (2011).
424 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).
425 Telschow, A., Engelstadter, J., Yamamura, N., Hammerstein, P. & Hurst, G. D. Asymmetric gene flow and constraints on adaptation caused by sex ratio distorters. *Journal of evolutionary biology* **19**, 869-878, doi:10.1111/j.1420-9101.2005.01049.x (2006).
426 Hoffmann, A. A. *et al.* Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. *Nature* **476**, 454-457, doi:10.1038/nature10356 (2011).
427 Magni, G. E. 'Sex-Ratio': a Non-Mendelian Character in Drosophila bifasciata. *Nature* **172**, 81, doi:10.1038/172081a0 (1953).
428 Harumoto, T. & Lemaitre, B. Male-killing toxin in a bacterial symbiont of Drosophila. *Nature* **557**, 252-255, doi:10.1038/s41586-018-0086-2 (2018).
429 Beckmann, J. F., Ronau, J. A. & Hochstrasser, M. A Wolbachia deubiquitylating enzyme induces cytoplasmic incompatibility. *Nature microbiology* **2**, 17007, doi:10.1038/nmicrobiol.2017.7 (2017).
430 Charlat, S. *et al.* Extraordinary flux in sex ratio. *Science (New York, N.Y.)* **317**, 214, doi:10.1126/science.1143369 (2007).
431 Sheeley, S. L. & McAllister, B. F. Mobile male-killer: similar Wolbachia strains kill males of divergent Drosophila hosts. *Heredity* **102**, 286-292, doi:10.1038/hdy.2008.126 (2009).
432 Bordenstein, S. R. & Wernegreen, J. J. Bacteriophage flux in endosymbionts (Wolbachia): infection frequency, lateral transfer, and recombination rates. *Molecular biology and evolution* **21**, 1981-1991, doi:10.1093/molbev/msh211 (2004).
433 Lindsey, A. R. I. *et al.* Evolutionary Genetics of Cytoplasmic Incompatibility Genes cifA and cifB in Prophage WO of Wolbachia. *Genome biology and evolution* **10**, 434-451, doi:10.1093/gbe/evy012 (2018).
434 Sampson, B. J. *et al.* 1133 edn 411-418 (International Society for Horticultural Science (ISHS), Leuven, Belgium).
435 Curtsinger, J. W. Components of selection in X chromosome lines of Drosophila melanogaster: sex ratio modification by meiotic drive and viability selection. *Genetics* **108**, 941-952 (1984).
436 Montenegro, H., Souza, W. N., da Silva Leite, D. & Klaczko, L. B. Male-killing selfish cytoplasmic element causes sex-ratio distortion in Drosophila melanogaster. *Heredity* **85 Pt 5**, 465-470 (2000).

437 Dyer, K. A. & Jaenike, J. Evolutionary dynamics of a spatially structured host-parasite association: Drosophila innubila and male-killing Wolbachia. *Evolution; international journal of organic evolution* **59**, 1518-1528 (2005).
438 Petrella, L. N., Smith-Leiker, T. & Cooley, L. The Ovhts polyprotein is cleaved to produce fusome and ring canal proteins required for Drosophila oogenesis. *Development (Cambridge, England)* **134**, 703-712, doi:10.1242/dev.02766 (2007).
439 Landmann, F., Orsi, G. A., Loppin, B. & Sullivan, W. Wolbachia-mediated cytoplasmic incompatibility is associated with impaired histone deposition in the male pronucleus. *PLoS pathogens* **5**, e1000343, doi:10.1371/journal.ppat.1000343 (2009).
440 Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10**, 845-858, doi:10.1038/nprot.2015.053 (2015).
441 Kim, M. *et al.* Noncanonical DNA-binding mode of repressor and its disassembly by antirepressor. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E2480-2488, doi:10.1073/pnas.1602618113 (2016).
442 Luscombe, N. M., Austin, S. E., Berman, H. M. & Thornton, J. M. An overview of the structures of protein-DNA complexes. *Genome biology* **1**, Reviews001, doi:10.1186/gb-2000-1-1-reviews001 (2000).
443 Pinto, S. B. *et al.* Transcriptional regulation of Culex pipiens mosquitoes by Wolbachia influences cytoplasmic incompatibility. *PLoS pathogens* **9**, e1003647 (2013).
444 Zhang, G., Hussain, M., O'Neill, S. L. & Asgari, S. Wolbachia uses a host microRNA to regulate transcripts of a methyltransferase, contributing to dengue virus inhibition in Aedes aegypti. *Proceedings of the National Academy of Sciences* **110**, 10276-10281 (2013).
445 Lux, S. A. *et al.* Effects of Irradiation on the Courtship Behavior of Medfly (Diptera, Tephritidae) Mass Reared for the Sterile Insect Technique. *The Florida Entomologist* **85**, 102-112 (2002).
446 Barclay, H. J. Modeling incomplete sterility in a sterile release program: interactions with other factors. *Population Ecology* **43**, 197-206, doi:10.1007/s10144-001-8183-7 (2001).
447 Telschow, A., Hammerstein, P. & Werren, J. H. The effect of Wolbachia versus genetic incompatibilities on reinforcement and speciation. *Evolution* **59**, 1607-1619 (2005).
448 Hurst, G. D. & McVean, G. A. T. Parasitic male-killing bacteria and the evolution of clutch size. *Ecological entomology* **23**, 350-353 (1998).
449 Baym, M. *et al.* Inexpensive multiplexed library preparation for megabase-sized genomes. *PloS one* **10**, e0128036, doi:10.1371/journal.pone.0128036 (2015).
450 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology* **19**, 455-477, doi:10.1089/cmb.2012.0021 (2012).
451 Longdon, B., Fabian, D. K., Hurst, G. D. & Jiggins, F. M. Male-killing Wolbachia do not protect Drosophila bifasciata against viral infection. *BMC microbiology* **12 Suppl 1**, S8, doi:10.1186/1471-2180-12-s1-s8 (2012).
452 Ellegaard, K. M., Klasson, L., Naslund, K., Bourtzis, K. & Andersson, S. G. Comparative genomics of Wolbachia and the bacterial species concept. *PLoS genetics* **9**, e1003381, doi:10.1371/journal.pgen.1003381 (2013).
453 Boetzer, M. & Pirovano, W. Toward almost closed genomes with GapFiller. *Genome biology* **13**, R56, doi:10.1186/gb-2012-13-6-r56 (2012).

454 Sullivan, W., Ashburner, M. & Hawley, R. S. *Drosophila protocols*. (Cold Spring Harbor Laboratory Press, 2000).
455 Hall, S. & Ward, R. E. t. Septate Junction Proteins Play Essential Roles in Morphogenesis Throughout Embryonic Development in Drosophila. *G3 (Bethesda, Md.)* **6**, 2375-2384, doi:10.1534/g3.116.031427 (2016).
456 Capra, J. A. & Singh, M. Predicting functionally important residues from sequence conservation. *Bioinformatics (Oxford, England)* **23**, 1875-1882, doi:10.1093/bioinformatics/btm270 (2007).
457 Perlmutter, J. I. & Bordenstein, S. R. Microorganisms in the reproductive tissues of arthropods. *Nature Reviews Microbiology*, 1-15 (2020).
458 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).
459 Waltz, E. US government approves' killer'mosquitoes to fight disease. *Nature News* (2017).
460 Dorigatti, I., McCormack, C., Nedjati-Gilani, G. & Ferguson, N. M. Using Wolbachia for dengue control: Insights from modelling. *Trends in parasitology* **34**, 102-113 (2018).
461 Zheng, X. *et al.* Incompatible and sterile insect techniques combined eliminate mosquitoes. *Nature* **572**, 56-61 (2019).
462 Cooper, B. S., Vanderpool, D., Conner, W. R., Matute, D. R. & Turelli, M. Wolbachia Acquisition by Drosophila yakuba-Clade Hosts and Transfer of Incompatibility Loci Between Distantly Related Wolbachia. *Genetics* **212**, 1399-1419 (2019).
463 Lindsey, A. R. *et al.* Evolutionary genetics of cytoplasmic incompatibility genes cifA and cifB in prophage WO of Wolbachia. *Genome biology and evolution* **10**, 434-451 (2018).
464 Yamada, R., Iturbe-Ormaetxe, I., Brownlie, J. & O'Neill, S. Functional test of the influence of Wolbachia genes on cytoplasmic incompatibility expression in Drosophila melanogaster. *Insect molecular biology* **20**, 75-85 (2011).
465 Hoffmann, A. A., Hercus, M. & Dagher, H. Population dynamics of the Wolbachia infection causing cytoplasmic incompatibility in Drosophila melanogaster. *Genetics* **148**, 221-231 (1998).
466 Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10**, 845 (2015).
467 Juillerat, A. *et al.* BurrH: a new modular DNA binding protein for genome engineering. *Scientific reports* **4**, 3831 (2014).
468 Meyer, D. F. *et al.* Searching algorithm for type IV secretion system effectors 1.0: a tool for predicting type IV effectors and exploring their genomic context. *Nucleic acids research* **41**, 9218-9229 (2013).
469 Noroy, C., Lefrançois, T. & Meyer, D. F. Searching algorithm for Type IV effector proteins (S4TE) 2.0: Improved tools for Type IV effector prediction, analysis and comparison in proteobacteria. *PLoS computational biology* **15**, e1006847 (2019).
470 Iturbe-Ormaetxe, I., Burke, G. R., Riegler, M. & O'Neill, S. L. Distribution, expression, and motif variability of ankyrin domain genes in Wolbachia pipientis. *Journal of bacteriology* **187**, 5136-5145 (2005).
471 Siozios, S. *et al.* The diversity and evolution of Wolbachia ankyrin repeat domain genes. *PLoS One* **8**, e55390 (2013).
472 Voth, D. E. *et al.* The Coxiella burnetii cryptic plasmid is enriched in genes encoding type IV secretion system substrates. *Journal of bacteriology* **193**, 1493-1503 (2011).

473 Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. SMART, a simple modular architecture research tool: identification of signaling domains. *Proceedings of the National Academy of Sciences* **95**, 5857-5864 (1998).
474 Zimmermann, L. *et al.* A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *Journal of molecular biology* **430**, 2237-2243 (2018).
475 Hecht, A. *et al.* Measurements of translation initiation from all 64 codons in E. coli. *Nucleic acids research* **45**, 3615-3626 (2017).
476 Bordenstein, S. R., Marshall, M. L., Fry, A. J., Kim, U. & Wernegreen, J. J. The tripartite associations between bacteriophage, Wolbachia, and arthropods. *PLoS pathogens* **2**, e43 (2006).
477 Fry, A., Palmer, M. & Rand, D. Variable fitness effects of Wolbachia infection in Drosophila melanogaster. *Heredity* **93**, 379 (2004).
478 Sinkins, S. P. *et al.* Wolbachia variability and host effects on crossing type in Culex mosquitoes. *Nature* **436**, 257 (2005).
479 Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z. & Blüthgen, N. Efficient translation initiation dictates codon usage at gene start. *Molecular systems biology* **9** (2013).
480 Layton, E. M., On, J., Perlmutter, J. I., Bordenstein, S. R. & Shropshire, J. D. Paternal Grandmother Age Affects the Strength of Wolbachia-Induced Cytoplasmic Incompatibility in Drosophila melanogaster. *mBio* **10** (2019).
481 Blagrove, M. S., Arias-Goeta, C., Failloux, A.-B. & Sinkins, S. P. Wolbachia strain wMel induces cytoplasmic incompatibility and blocks dengue transmission in Aedes albopictus. *Proceedings of the National Academy of Sciences* **109**, 255-260 (2012).
482 Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics (Oxford, England)* **27**, 1164-1165 (2011).
483 Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology* **52**, 696-704 (2003).
484 Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)* **17**, 754-755 (2001).
485 Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)* **19**, 1572-1574 (2003).
486 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic acids research* (2019).
487 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).
488 Hamm, C. A. *et al.* Wolbachia do not live by reproductive manipulation alone: infection polymorphism in Drosophila suzukii and D. subpulchrella. *Molecular ecology* **23**, 4871-4885 (2014).
489 Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The vienna RNA websuite. *Nucleic acids research* **36**, W70-W74 (2008).
490 Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms for molecular biology* **6**, 26 (2011).
491 Jiggins, F. M., Randerson, J. P., Hurst, G. D. & Majerus, M. E. How can sex ratio distorters reach extreme prevalences? Male-killing Wolbachia are not suppressed and have near-perfect vertical transmission efficiency in Acraea encedon. *Evolution* **56**, 2290-2295 (2002).
492 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome research* **12**, 996-1006 (2002).

493 Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic acids research* **31**, 3429-3431 (2003).
494 Malogolowkin, C. & Poulson, D. F. Infective transfer of maternally inherited abnormal sex-ratio in Drosophila willistoni. *Science (New York, N.Y.)* **126**, 32 (1957).
495 Berec, L., Maxin, D., & Bernhauerova, V. Male-killing bacteria as agents of insect pest control. *J Appl Ecol* **53**, 1270-1279 (2016).
496 Yamada, R., Iturbe-Ormaetxe, I., Brownlie, J. C. & O'Neill, S. L. Functional test of the influence of Wolbachia genes on cytoplasmic incompatibility expression in Drosophila melanogaster. *Insect molecular biology* **20**, 75-85, doi:10.1111/j.1365-2583.2010.01042.x (2011).
497 Harumoto, T. & Lemaitre, B. Male-killing toxin in a bacterial symbiont of Drosophila. *Nature*, doi:10.1038/s41586-018-0086-2 (2018).
498 Jernigan, K. K. & Bordenstein, S. R. Ankyrin domains across the Tree of Life. *PeerJ* **2**, e264, doi:10.7717/peerj.264 (2014).
499 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. N. Male-killing Wolbachia in two species of insect. *Proceedings of the Royal Society of London B: Biological Sciences* **266**, 735-740 (1999).

Appendix A. Chapter III Supplementary Information

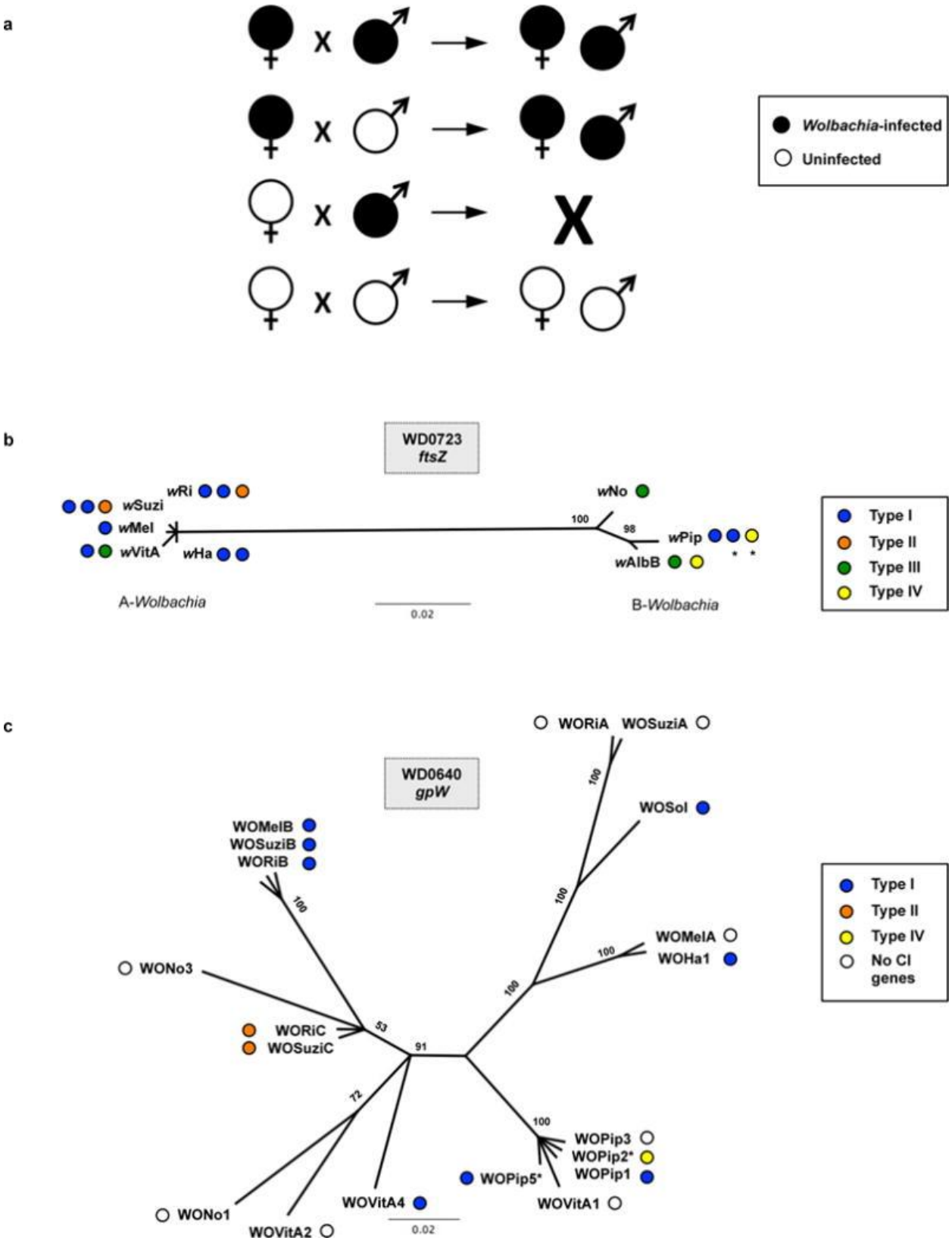


Figure A-1. CI and the evolution of *Wolbachia* and prophage WO genes.
(a) Diagram shows the effect of parental *Wolbachia* infection on progeny viability and infection status. CI (embryonic inviability) occurs in crosses between *Wolbachia*-infected males and uninfected females. *Wolbachia*-infected females mated to infected males rescue the inviability. (b) Bayesian phylogenies based on a 393-aa alignment of WD0723, the *w*Mel *ftsZ* gene, and its homologs and (c) a 70-aa alignment of WD0640, the phage WO *gpW* gene, and its homologs. Trees are based on JTT+G and CpRev+I models of evolution, respectively, and are unrooted. Consensus support values are shown at the nodes. (*) indicates that the CI genes are not included in Fig. 3-1. The WOPip5 homolog is truncated while the WOPip2 and second *w*AlbB homologs are highly divergent from WD0632.

Figure A-2. WD0631/WD0632 homologs always associate with the eukaryotic association module in prophage WO regions.

CI gene homologs are labeled and colored pink. Structural modules are labeled as baseplate, head or tail. The WD0611-WD0621 label highlights a conserved gene cluster that is often associated

with the CI genes. Only one phage haplotype is shown per *Wolbachia* strain when multiple copies of the same type are present.

a

WD0631-like % aa identity

	WORiB	WORiC	WOHa1	*w*No
WOMelB	99	46	67	31
WORiB		46	68	31
WORiC			44	33
WOHa1				31

WD0632-like % aa identity

	WORiB	WORiC	WOHa1	*w*No
WOMelB	99	30	62	29
WORiB		30	62	29
WORiC			31	36
WOHa1				30

b

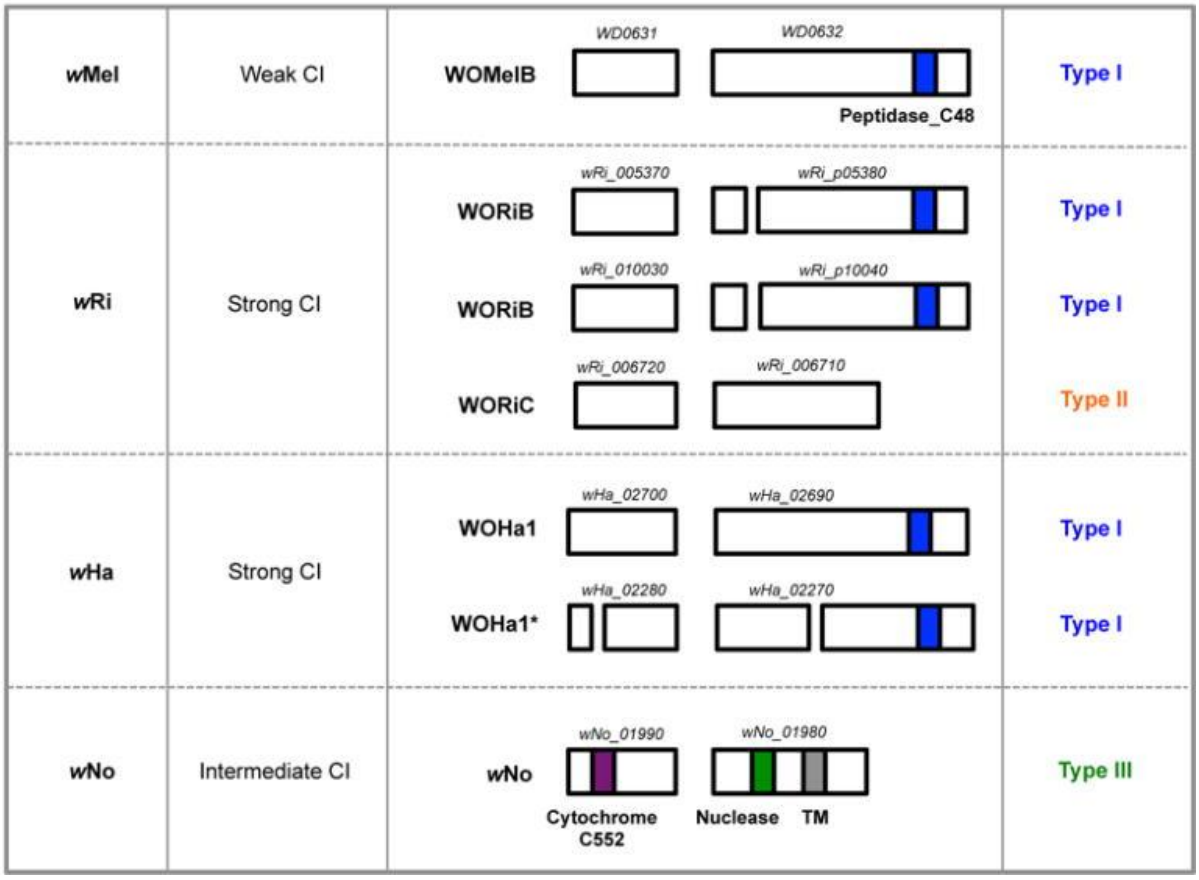


Figure A-3. *Wolbachia* CI patterns correlate with WD0631/WD0632 homolog similarity and copy number.

(a) The % amino acid (aa) identity between each WD0631/WD0632 homolog correlates with *Wolbachia* compatibility patterns. The only compatible cross, *w*Mel males × *w*Ri females, features close homology between WOMelB and WORiB. All other crosses are greater than 30% divergent and are bidirectionally incompatible. Each "% aa identity" value is based on the region of query coverage in a 1:1 BLASTp analysis. (b) CI strength, protein architecture and clade type are listed for each of the *Wolbachia* strains shown in Figure 1d. (*) indicates the proteins are disrupted and not included in comparison analyses.

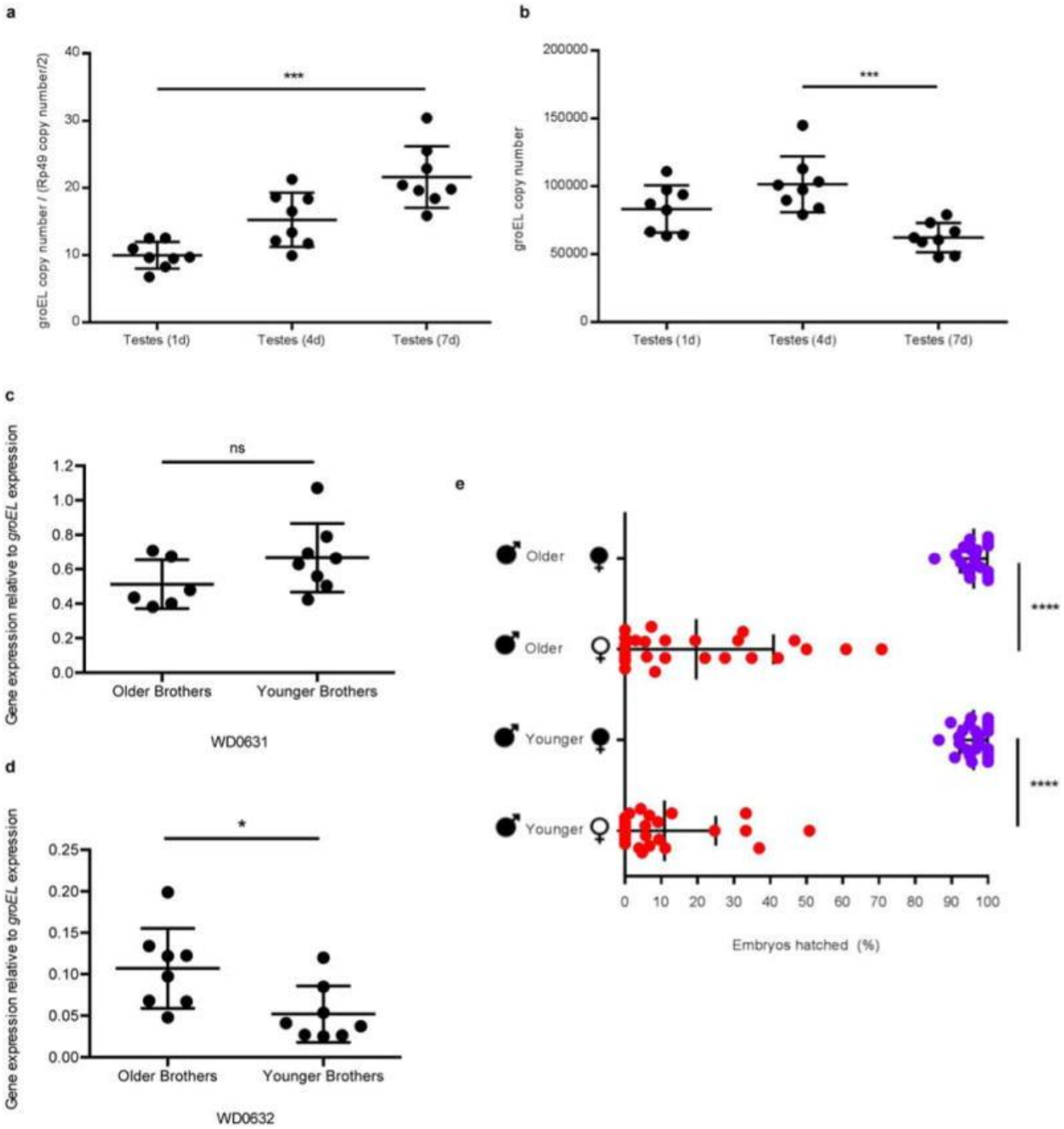


Figure A-4. *Wolbachia* titers, the male age effect, and the younger brother effect.
(a) Relative *Wolbachia* titers in wild type lines do not decrease with age. DNA copy number of the *w*Mel *groEL* gene is shown normalized to *D. melanogaster rp49* gene copy number in testes at the indicated ages. (b) Absolute *Wolbachia* titers do not decrease with male age. (c, d) In *w*Mel-infected males, WD0631 gene expression is equal between older (first day of emergence) and younger brothers (fifth day of emergence) while WD0632 gene expression is slightly higher in early emerging brothers. (e) There is no statistical difference in CI penetrance between older and younger brothers. Bars show mean and standard deviation. * = P<0.05, *** = P<0.001, **** = P<0.0001 by ANOVA with Kruskal-Wallis test and Dunn's multiple test correction for (a), (b),

and (e), and two-tailed Mann-Whitney U test used for (c), and (d). Exact p-values are provided in Table A-7. These experiments have been performed once.

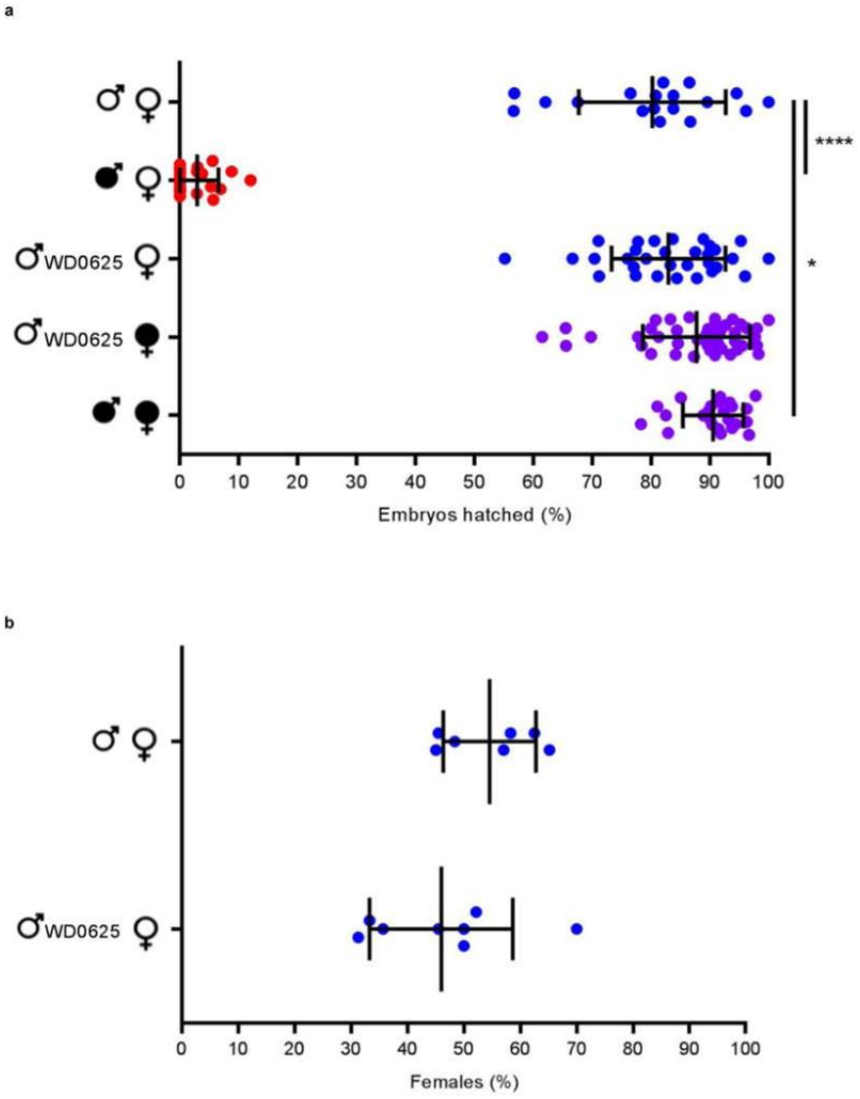


Figure A-5. WD0625 transgene expression does not induce CI-like defects.
Expression of control gene WD0625 in one-day-old uninfected males does not affect (a) embryo hatch rates or (b) sex ratios. Infection status is designated with filled in symbols for a *w*Mel-infected parent or open symbols for an uninfected parent. Transgenic flies are labeled with their transgene to the right of their male/female symbol. Unlabeled symbols represent wild type flies. Data points are colored according to the type of cross, with blue indicating no CI, red indicating a CI cross, and purple indicating a rescue cross with *w*Mel-infected females. Bars indicate mean and standard deviation. * = P<0.05, *** = P<0.001 by ANOVA with Kruskal-Wallis test and Dunn's multiple test correction. Exact p-values are provided in Table A-7. This experiment has been replicated three times.

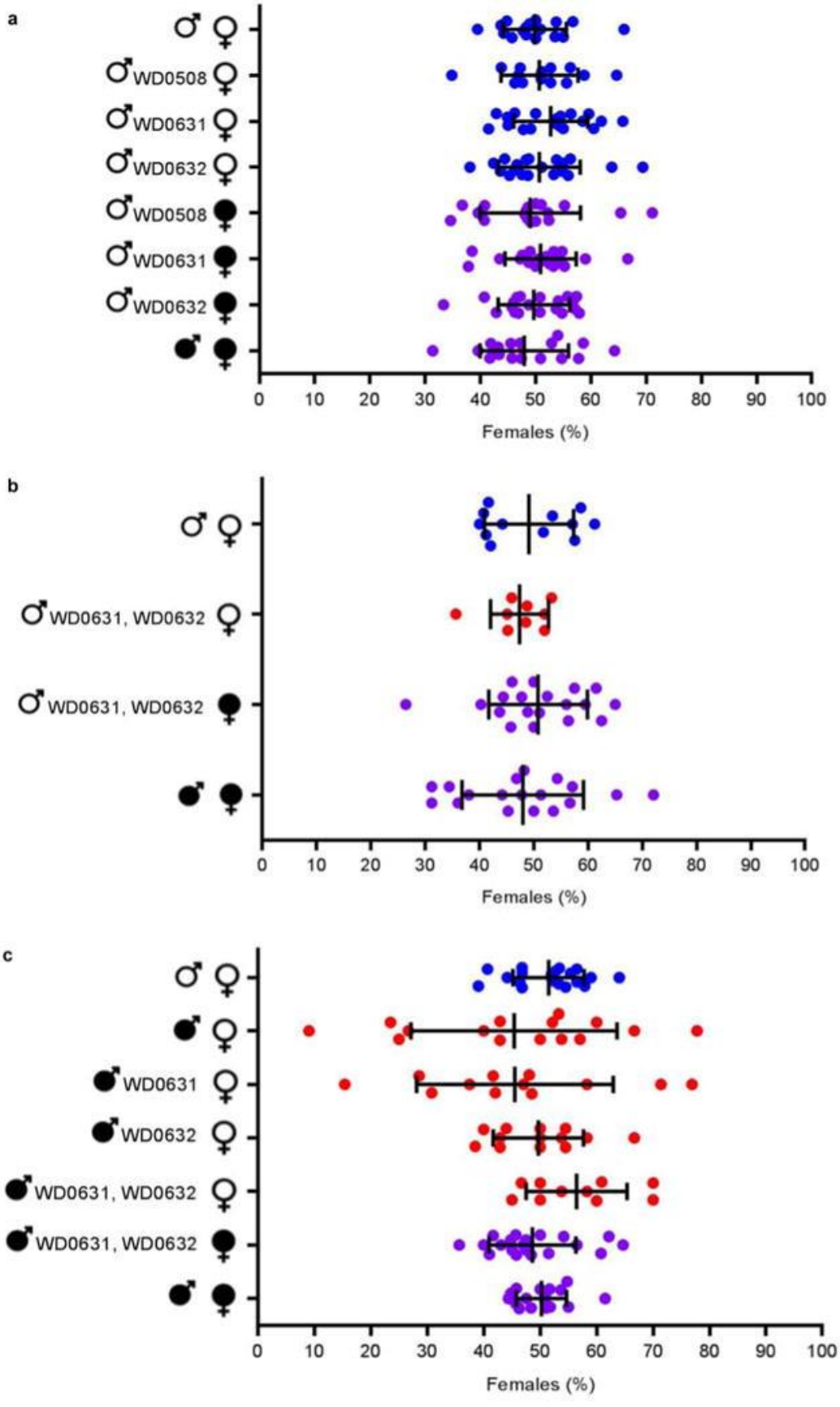

a
♂ ♀
♂WD0508 ♀
♂WD0631 ♀
♂WD0632 ♀
♂WD0508 ♂
♂WD0631 ♂
♂WD0632 ♂
♂ ♂
0 10 20 30 40 50 60 70 80 90 100
Females (%)
b
♂ ♀
♂WD0631, WD0632 ♀
♂WD0631, WD0632 ♂
♂ ♂
0 10 20 30 40 50 60 70 80 90 100
Females (%)
c
♂ ♀
♂ ♀
♂WD0631 ♀
♂WD0632 ♀
♂WD0631, WD0632 ♀
♂WD0631, WD0632 ♂
♂ ♂
0 10 20 30 40 50 60 70 80 90 100
Females (%)

Figure A-6. Expression of transgenes does not alter sex ratios.
Graphs correspond to the same crosses as Fig. 3. Infection status is designated with filled in symbols for a *w*Mel-infected parent or open symbols for an uninfected parent. Transgenic flies are labeled with their transgene to the right of their gender symbol. Unlabeled gender symbols represent WT flies. Data points are colored according to the type of cross, with blue indicating no CI, red indicating a CI cross, and purple indicating a rescue cross with *w*Mel-infected females. Bars indicate mean and standard deviation. Statistics include a Kruskal-Wallis tests and Dunn's multiple test corrections. Figs. A-6a and A-6c have been performed once, while Fig. A-6b has been performed twice.

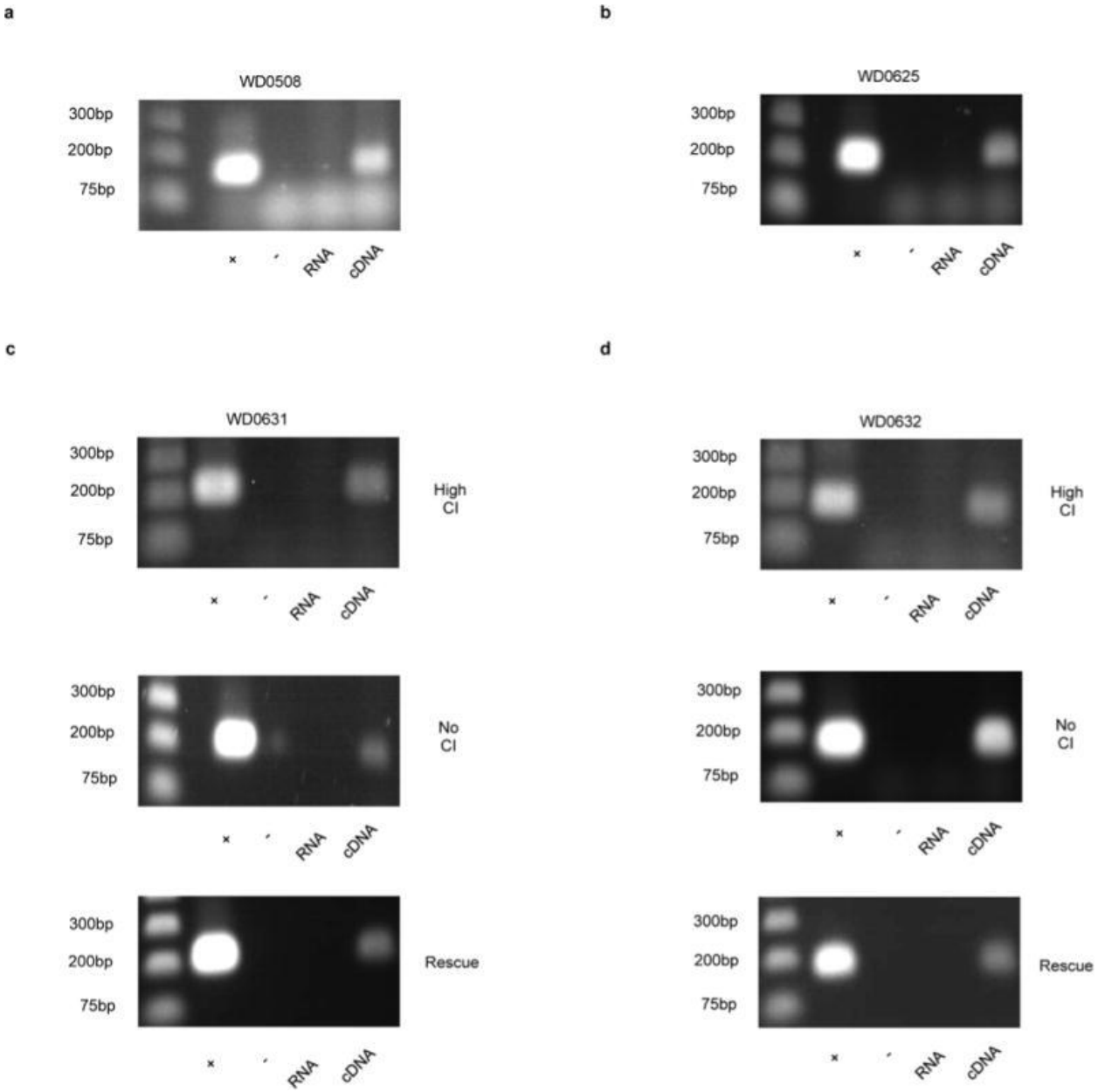


Figure A-7. Transgenes are expressed in testes.
(a, b) WD0508 and WD0625 transgenes are expressed in testes as evident by PCR performed against cDNA generated from dissected males utilized in Fig. 3a and Fig. A-5a, respectively. (c, d) WD0631 and WD0632 transgenes are expressed in the testes from transgenic males specifically inducing high CI, no CI, or rescued CI. Testes were removed from males used in a replicate of Fig. 3b. This experiment has been performed once.

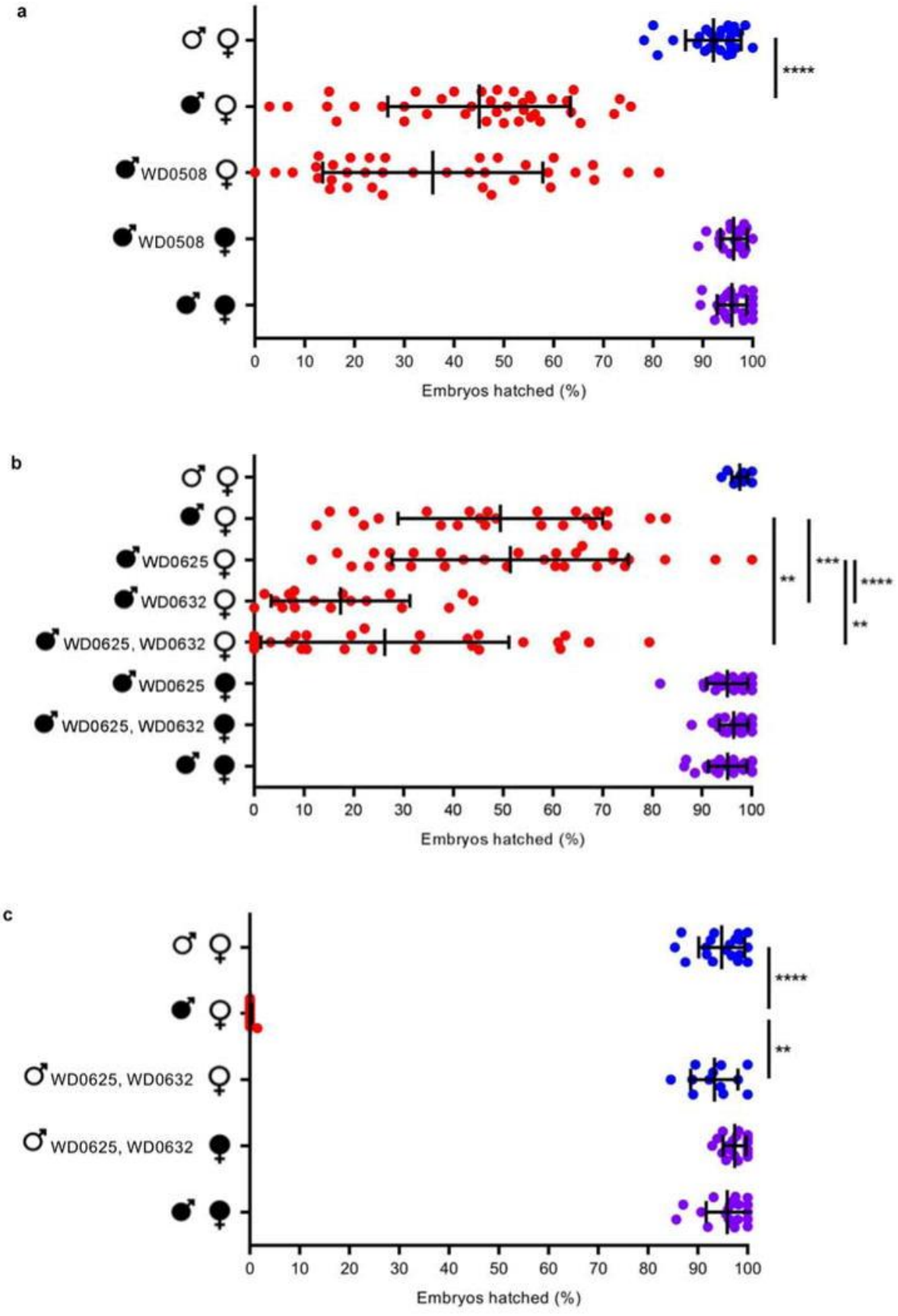

a
WD0508
WD0508
****
0 10 20 30 40 50 60 70 80 90 100
Embryos hatched (%)
b
WD0625
WD0632
WD0625, WD0632
WD0625
WD0625, WD0632
**
***
****
**
0 10 20 30 40 50 60 70 80 90 100
Embryos hatched (%)
c
WD0625, WD0632
WD0625, WD0632
****
**
0 10 20 30 40 50 60 70 80 90 100
Embryos hatched (%)

Figure A-8. Transgenic expression of WD0508, WD0625, and WD0625/WD0632 does not enhance or induce CI.
(a) The WD0508 transgene alone does not enhance CI in two- to four-day-old infected males. (b) The WD0625 transgene alone does not enhance CI either; conversely, WD0632 does enhance CI as previously shown in Fig. 3C. The WD0625 transgene together with WD0632 does not enhance CI further than WD0632 alone. (**c**) WD0625/WD0632 dual expression cannot induce CI in uninfected one-day-old males. Infection status is designated with filled in symbols for a *w*Mel-infected parent or open symbols for an uninfected parent. Transgenic flies are labeled with their transgene to the right of their male/female symbol. Unlabeled symbols represent WT flies. Data points are colored according to the type of cross, with blue indicating no CI, red indicating a CI cross, and purple indicating a rescue cross with *w*Mel-infected females. Bars indicate mean and standard deviation. ** = P<0.01, *** = P<0.001, **** = P<0.0001 by ANOVA with Kruskal-Wallis test and Dunn's multiple test correction. Exact p-values are provided in Table A-7. These experiments have been done twice (A-8a, 8c), three times (A-8b, WD0625, WD0632), or once (A-8b, WD0625/WD0632).

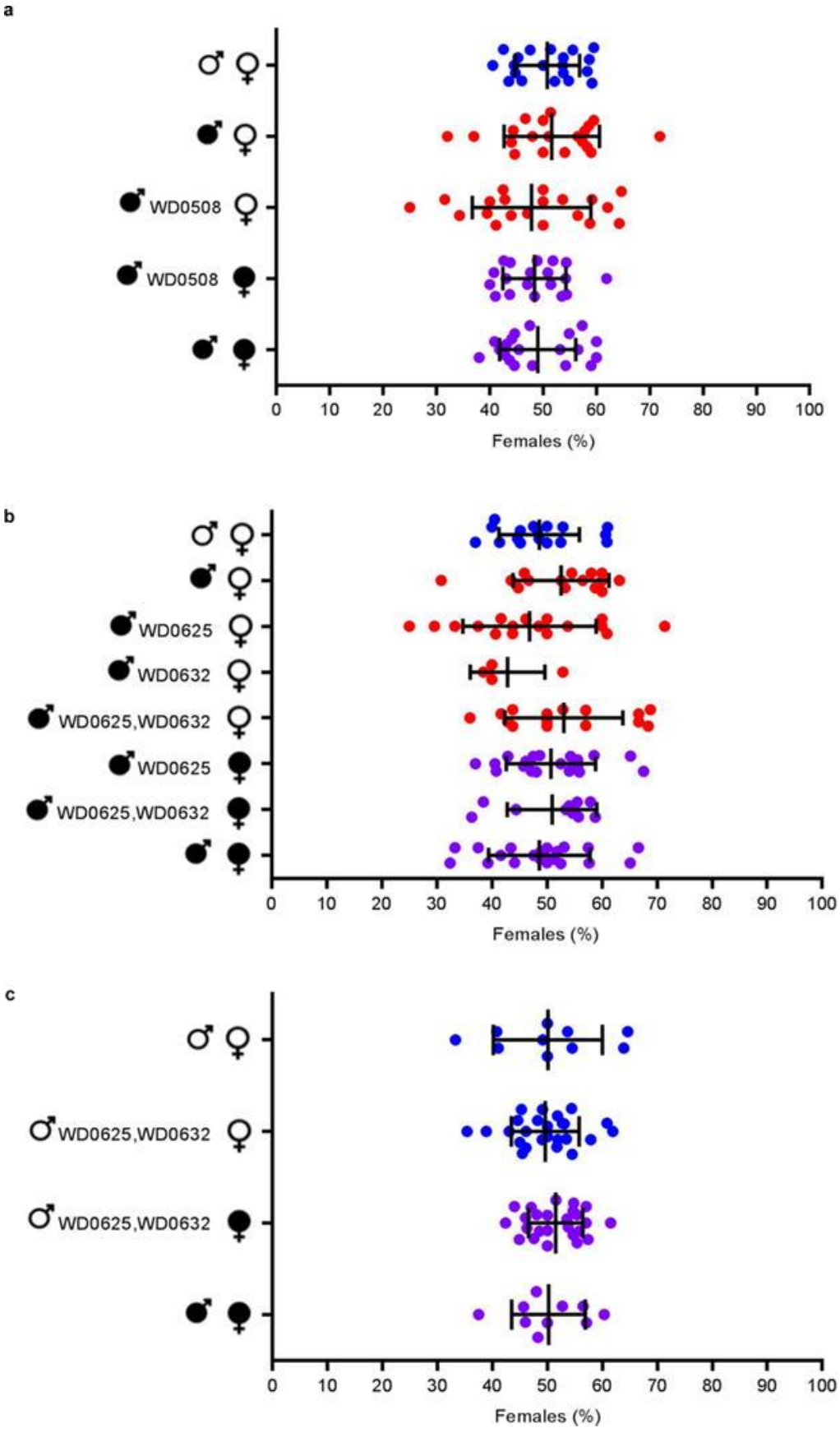

a
♂ ♀
♂ ♀
♂ WD0508 ♀
♂ WD0508 ♀
♂ ♀
0 10 20 30 40 50 60 70 80 90 100
Females (%)
b
♂ ♀
♂ ♀
♂ WD0625 ♀
♂ WD0632 ♀
♂ WD0625,WD0632 ♀
♂ WD0625 ♀
♂ WD0625,WD0632 ♀
♂ ♀
0 10 20 30 40 50 60 70 80 90 100
Females (%)
c
♂ ♀
♂ WD0625,WD0632 ♀
♂ WD0625,WD0632 ♀
♂ ♀
0 10 20 30 40 50 60 70 80 90 100
Females (%)

Figure A-9. Transgenic expression of control genes does not affect sex ratios.
All flies are from same crosses shown in Fig. A-8, except for Fig. A-9c which comes from a replicate experiment. Infection status is designated with filled in symbols for a *w*Mel-infected parent or open symbols for an uninfected parent. Transgenic flies are labeled with their transgene to the right of their male/female symbol. Unlabeled symbols represent WT flies. Data points are colored according to the type of cross, with blue indicating no CI, red indicating a CI cross, and purple indicating a rescue cross with *w*Mel-infected females. Bars indicate mean and standard deviation. Statistics performed by ANOVA with Kruskal-Wallis test and Dunn's multiple test correction. These experiments have been done twice (A-9b) or once (A-9a,c).

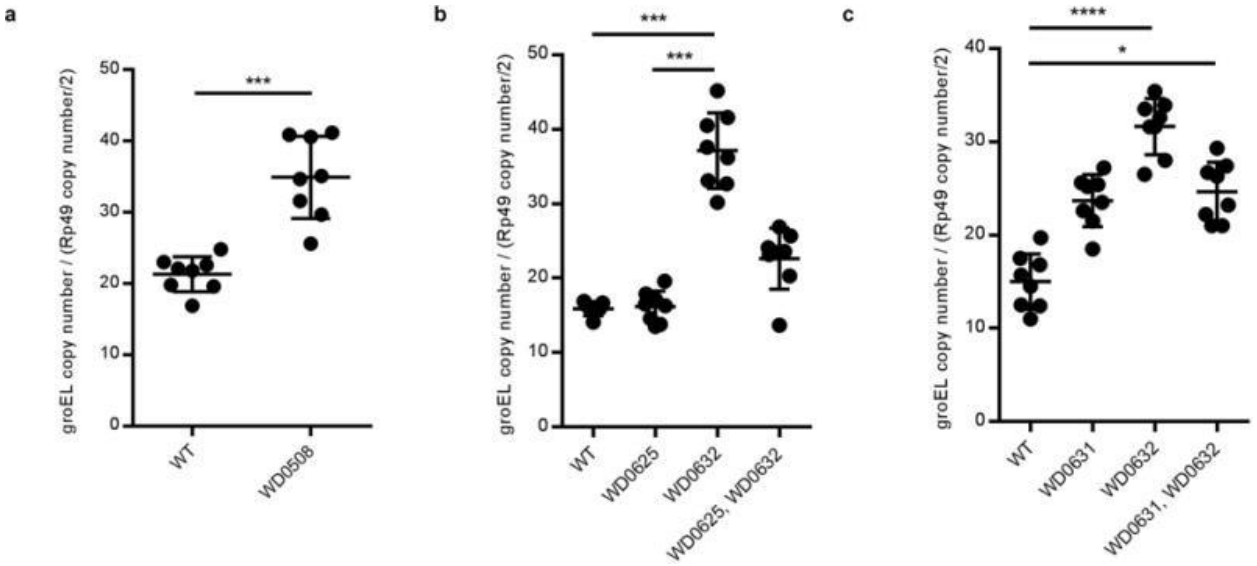


Figure A-10. There is variation in *Wolbachia* titers in transgenic lines.
(a–c) Relative *Wolbachia* titers are higher in WD0508, WD0631, and WD0632 transgenic lines in comparison to wild type lines. This does not occur in the WD0625 transgenic line nor does there appear to be an additive effect. DNA copy number of *w*Mel *groEL* gene is shown normalized to *D. melanogaster rp49* gene copy number in testes of the indicated strains. Bars show mean and standard deviation. * = P<0.05, *** = P<0.001, **** = P<0.0001 for two-tailed Mann-Whitney U test (a) and Kruskal-Wallis test with Dunn's multiple test correction (b–c). Exact p-values are provided in Table A-7. These experiments have been done once.

Table A-1. Core CI genome. This table shows the 161 unique *w*Mel genes identified as the "Core CI Genome" in Figure 1a.

***w* Mel Locus Tag**	**Gene Description**
WD0035	ankyrin repeat-containing protein
WD0038	Protein tolB
WD0046	reverse transcriptase, interruption-N
WD0049	hypothetical protein
WD0056	major facilitator family transporter

WD0061	hypothetical protein
WD0064	Pyridoxine 5'-phosphate synthase
WD0069	hypothetical protein
WD0074	hypothetical protein
WD0077	hypothetical protein
WD0078	hypothetical protein
WD0079	hypothetical protein
WD0092	DNA processing chain A
WD0099	multidrug resistance protein
WD0100	sugE protein
WD0131	hypothetical protein
WD0139	TenA family transcription regulator
WD0140	TenA family transcription regulator
WD0168	major facilitator family transporter
WD0200	hypothetical protein
WD0208	hypothetical protein
WD0211	hypothetical protein
WD0214	hypothetical protein
WD0217	phage uncharacterized protein
WD0231	hypothetical protein
WD0234	hypothetical protein
WD0240	transposase, IS5 family, degenerate
WD0255	transcriptional regulator, putative
WD0257	DNA repair protein RadC, truncation
WD0258	Reverse transcriptase, frame shift
WD0274	hypothetical protein
WD0278	Prophage LambdaW1, minor tail protein Z
WD0279	hypothetical protein
WD0281	hypothetical protein
WD0282	prophage LambdaW1, baseplate assembly protein W, putative
WD0283	prophage LambdaW1, baseplate assembly protein J, putative
WD0284	hypothetical protein
WD0285	Prophage LambdaW1, ankyrin repeat protein
WD0286	ankyrin repeat-containing prophage LambdaW1

WD0288	prophage LambdaW1, site-specific recombinase resolvase family protein
WD0315	hypothetical protein
WD0324	hypothetical protein
WD0329	transposase, IS5 family, degenerate
WD0336	transposase, IS5 family, degenerate
WD0338	hypothetical protein
WD0345	RND family efflux transporter MFP subunit
WD0376	potassium uptake protein TrKH, frame shift
WD0382	hypothetical protein
WD0385	ankyrin repeat-containing protein
WD0396	reverse transcriptase, truncation
WD0407	Na+/H+ antiporter, putative
WD0426	hypothetical protein
WD0431	glycosyl transferase, group 2 family protein
WD0447	phage prohead protease
WD0458	HK97 family phage major capsid protein
WD0472	AAA family ATPase
WD0480	hypothetical protein
WD0481	hypothetical protein
WD0482	SPFH domain-containing protein/band 7 family protein
WD0483	M23/M37 peptidase domain-containing protein
WD0498	ankyrin repeat-containing protein
WD0501	surface antigen-related protein
WD0506	Reverse transcriptase, frame shift
WD0507	DNA repair protein RadC, truncation
WD0508	transcriptional regulator, putative
WD0515	reverse transcriptase, interruption-C
WD0518	reverse transcriptase, interruption-N
WD0538	reverse transcriptase, truncation
WD0604	hypothetical protein
WD0606	Reverse transcriptase, frame shift
WD0623	transcriptional regulator, putative
WD0624	conserved domain protein, frame shift
WD0625	DNA repair protein RadC, putative

WD0626	transcriptional regulator, putative
WD0628	hypothetical protein
WD0631 WD0632	hypothetical protein hypothetical protein (Ulp1/Peptidase_C48)
WD0633	Prophage LambdaW5, ankyrin repeat domain protein
WD0634	prophage LambdaW5, site-specific recombinase resolvase family protein
WD0636	ankyrin repeat-containing prophage LambdaW1
WD0638	hypothetical protein
WD0639	prophage LambdaW5, baseplate assembly protein J, putative
WD0640	prophage LambdaW5, baseplate assembly protein W, putative
WD0641	hypothetical protein
WD0642	prophage LambdaW5, baseplate assembly protein V
WD0643	hypothetical protein
WD0644	Prophage LambdaW5, minor tail protein Z
WD0645	reverse transcriptase, truncation
WD0686	hypothetical protein
WD0693	reverse transcriptase, putative
WD0696	hypothetical protein
WD0702	hypothetical protein
WD0713	hypothetical protein
WD0718	conserved hypothetical protein, truncated
WD0721	Mg chelatase-related protein
WD0724	hypothetical protein
WD0730	phosphatidylglycerophosphatase A, putative
WD0733	hypothetical protein
WD0748	hypothetical protein
WD0749	transposase, IS5 family, degenerate
WD0750	PQQ repeat-containing protein
WD0764	hypothetical protein
WD0787	araM protein
WD0790	hypothetical protein
WD0818	hypothetical protein
WD0823	hypothetical protein
WD0826	hypothetical protein

WD0834	conserved hypothetical protein, degenerate
WD0835	hypothetical protein
WD0874	transposase, truncated
WD0875	IS5 family transposase
WD0880	coenzyme PQQ synthesis protein C, putative
WD0882	FolK
WD0883	dihydropteroate synthase, putative
WD0884	dihydrofolate reductase
WD0887	DNA repair protein RadA
WD0901	transposase, IS110 family, degenerate
WD0903	transposase, IS5 family, degenerate
WD0908	transposase, degenerate
WD0911	transposase, IS5 family, degenerate
WD0914	hypothetical protein
WD0932	IS5 family transposase
WD0935	transposase, IS5 family, interruption-C
WD0941	transposase, degenerate
WD0947	IS5 family transposase
WD0958	hypothetical protein
WD0964	hypothetical protein
WD0975	hypothetical protein
WD0995	reverse transcriptase
WD0999	hypothetical protein
WD1002	hypothetical protein
WD1012	HK97 family phage portal protein
WD1015	hypothetical protein
WD1016	phage uncharacterized protein
WD1041	surface protein-related protein
WD1047	sodium/alanine symporter family protein
WD1052	folylpolyglutamate synthase
WD1069	hypothetical protein
WD1073	N-acetylmuramoyl-L-alanine amidase
WD1091	tRNA (guanine-N(7)-)-methyltransferase
WD1114	LipB
WD1118	hypothetical protein

WD1126	hypothetical protein
WD1131	conserved hypothetical protein, degenerate
WD1132	phage uncharacterized protein
WD1138	reverse transcriptase, putative
WD1159	Pyridoxine/pyridoxamine 5'-phosphate oxidase
WD1160	ComEC/Rec2 family protein
WD1161	hypothetical protein
WD1162	ribosomal large subunit pseudouridine synthase D
WD1163	diacylglycerol kinase
WD1175	hypothetical protein
WD1179	hypothetical protein
WD1204	TPR domain-containing protein
WD1212	16S ribosomal RNA methyltransferase RsmE
WD1218	ParB family protein
WD1242	hypothetical protein
WD1272	hypothetical protein
WD1310	hypothetical protein
WD1320	multidrug resistance protein D
WD1321	hypothetical protein

Table A-2. Genes divergent in *w*Au. *w*Mel genes that are absent or divergent in *w*Au as identified in Fig. 1b.

***w* Mel Locus Tag**	**Gene Description**
WD0019	transcription antitermination protein NusG, putative
WD0022	ribosomal protein L10
WD0034	PAZ Zwille/Arganaut/Piwi/ SiRNA binding domain
WD0072	hypothetical protein
WD0205	hypothetical protein
WD0244	hypothetical protein
WD0255	transcriptional regulator, putative
WD0256	hypothetical protein
WD0257	DNA repair protein RadC, truncation
WD0289	hypothetical protein
WD0297	hypothetical protein

WD0311	hypothetical protein
WD0320	trigger factor, putative
WD0349	hypothetical protein
WD0363	hypothetical protein
WD0366	hypothetical protein
WD0367	hypothetical protein
WD0369	hypothetical protein
WD0389	conserved hypothetical protein
WD0424	hypothetical protein
WD0449	hypothetical protein
WD0508	transcriptional regulator, putative
WD0512	ankyrin repeat domain protein
WD0553	hypothetical protein
WD0576	hypothetical protein
WD0577	hypothetical protein
WD0578	hypothetical protein
WD0579	hypothetical protein
WD0598	hypothetical protein
WD0607	hypothetical protein
WD0623	transcriptional regulator, putative
WD0624	conserved domain protein, authentic frameshift
WD0625	DNA repair protein RadC, putative
WD0626	transcriptional regulator, putative
WD0631 WD0632	hypothetical protein hypothetical protein (Ulp1/Peptidase_C48)
WD0633	prophage LambdaWp5, ankyrin repeat domain protein
WD0704	hypothetical protein
WD0723	cell division protein FtsZ
WD0746	hypothetical protein
WD0747	hypothetical protein
WD0806	hypothetical protein
WD0808	hypothetical protein
WD0809	hypothetical protein
WD0836	hypothetical protein
WD0837	hypothetical protein

WD0840	hypothetical protein
WD0850	rpsU-divergently transcribed protein
WD0854	membrane protein, putative
WD0877	hypothetical protein
WD0940	hypothetical protein
WD0946	hypothetical protein
WD0971	hypothetical protein
WD1038	hypothetical protein
WD1151	citrate synthase
WD1260	hypothetical protein
WD1287	hypothetical protein
WD1291	hypothetical protein
WD1311	Glycoside hydrolase 24
WD1313	conserved domain protein

Table A-3. *w*VitA transcriptome. *w*Mel homologs that are expressed (as measured by RNA) in the ovaries of *w*VitA-infected *Nasonia vitripennis*.

					Three individual replicates used for calculating the average		
***w* VitA Feature ID**	***w* Mel Locus Tag**	**Gene Description**	**12.1 wVitA - Mean RPKM**	**12.1 average # of reads mapped to gene**	**12.1 wVitA - 12_1_2M trimmed RNA-Seq - Total gene reads**	**12.1 wVitA - 12_1_3M RNA-Seq - Total gene reads**	**12.1 wVitA - 12_1_1M trimmed RNA-Seq - Total gene reads**
gwv_835	EXCLUDED	16S rRNA	351983.2481	66816.0	148928	22578	28942
gwv_1180	EXCLUDED	23S rRNA	157771.9274	53996.0	118025	18597	25366
gwv_528	WD1063	outer surface protein	835.3599645	33.0	19	47	33
gwv_664	WD0838	hypothetical protein	216.8355074	26.3	14	35	30
gwv_400	WD0745	putative outer membrane protein	205.5588118	24.0	17	35	20
rpoC	WD0024	DNA-directed RNA polymerase	91.44973092	20.7	9	38	15
gwv_424	WD0722	hypothetical protein	723.0305001	19.0	11	25	21
gwv_788	WD0906	S1 RNA binding domain protein	122.5244426	14.3	4	27	12
groEL	WD0307	chaperonin GroEL	140.8718458	13.7	10	23	8
gwv_846	WD1236	DNA/RNA helicase	183.4533269	13.0	8	20	11
gwv_141	WD0632	hypothetical protein (Ulp1/Peptidase_C48)	45.47043156	11.3	4	16	14
gwv_138	WD0292	ankryin repeat protein	52.0888594	10.3	3	13	15
fusA	WD0016	translation elongation factor G	81.2935939	9.7	5	13	11
gwv_1314	WD0950	uncharacterised protein family UPF0005	207.2377169	8.3	4	13	8
gwv_1093	WD0147	tetratricopeptide repeat family protein	17.79714113	7.3	4	13	5
gwv_142	WD0631	hypothetical protein	71.02829671	7.0	7	8	6
gyrA	WD1202	DNA gyrase	46.24634563	6.7	3	13	4
dnaK	WD0928	chaperone protein DnaK	55.20689465	6.3	3	5	11
gwv_968	WD0039	metallo-beta-lactamase superfamily protein	62.20883237	6.3	5	8	6
gwv_219	WD0550	ankryin repeat protein	104.550262	6.0	4	6	8

gwv_874	WD1249	sodium/hydrogen exchanger family protein	63.25273067	5.7	2	11	4
gwv_971	WD0041	putative membrane protein	98.05217219	5.7	4	7	6
gwv_889	WD1278	hypothetical protein	40.99214546	5.7	1	11	5
gwv_294	WD1071	cytochrome b	68.14464273	5.3	5	8	3
gwv_726	WD1064	RNA polymerase sigma-32 factor	103.3470918	5.0	3	8	4
gwv_848	WD1238	fructose-bisphosphate aldolase class 1	97.14626625	5.0	3	8	4
gwv_127	WD0337	hypothetical protein	70.90369473	5.0	3	3	9
rpsD			140.1432593	4.7	2	8	4
gwv_193			51.53566326	4.7	2	7	5
lpdA (gwv.assembly.1 388700..390089)			47.34146102	4.7	5	4	5
ftsZ			63.02213521	4.3	2	5	6
clpX			61.55998273	4.3	2	7	4
gwv_592			49.87489713	4.3	0	7	6
gwv_1267			228.5442649	4.0	3	6	3
gwv_1351			98.71359598	4.0	1	6	5
gwv_134			87.53953909	4.0	1	8	3
rpoA			68.14353225	4.0	2	7	3
agcS (gwv.assembly.1 617423..618869)			45.85405828	4.0	2	3	7
typA			38.00061768	4.0	2	5	5
lon			28.79549911	4.0	2	5	5
gwv_27			65.63670247	3.7	2	4	5
gwv_865			60.59688572	3.7	2	6	3
iscS			51.15879624	3.7	3	8	0
sucC			51.03343902	3.7	2	2	7
gwv_660			47.49917482	3.7	1	7	3
dnaX			41.74839386	3.7	2	7	2
clpB			29.67965234	3.7	0	9	2
gwv_46			22.57157345	3.7	2	4	5
gyrB			21.39989203	3.7	4	3	4
gwv_603			184.7219284	3.3	0	5	5
gwv_275			75.59102504	3.3	3	4	3
nusG			70.68705809	3.3	1	3	6
gwv_734			67.32779895	3.3	0	7	3
rho			52.5016369	3.3	0	6	4
gwv_837			46.44631685	3.3	1	2	7
gwv_1163			36.66487753	3.3	3	4	3
hslU			32.09423701	3.3	5	4	1
rplL			144.584441	3.0	1	6	2
rpoZ			123.1224335	3.0	2	4	3
gwv_868			114.347465	3.0	1	4	4
rplQ			109.5519599	3.0	2	2	5
gwv_38			77.90440588	3.0	1	2	6
gwv_549			77.50677716	3.0	2	5	2
rpsB			66.31737504	3.0	0	2	7
gwv_361			64.9017479	3.0	2	4	3
gwv_407			44.42206307	3.0	1	5	3
gwv_631			44.21350879	3.0	1	5	3
lysS			37.00384036	3.0	1	5	3
atpA			34.54520243	3.0	1	3	5
ctaD			32.3881063	3.0	2	4	3

sdhA			26.91908739	3.0	2	3	4
gwv_333			26.0913448	3.0	0	6	3
gwv_1345			25.21412951	3.0	1	5	3
gwv_560			169.7933214	2.7	1	3	4
gwv_263			83.02918988	2.7	2	4	2
rplA			79.54918497	2.7	0	3	5
gwv_417			72.92232064	2.7	0	6	2
gwv_517			61.11517361	2.7	2	3	3
sucD			55.92513437	2.7	1	4	3
atpB			54.38477172	2.7	2	1	5
gwv_774			53.70391551	2.7	0	5	3
trpS			50.65915791	2.7	1	5	2
dnaN			47.36465385	2.7	1	2	5
dnaA			43.46751424	2.7	0	8	0
hflK			41.27488537	2.7	2	3	3
gwv_293			36.43743379	2.7	2	2	4
gwv_390			35.89644056	2.7	1	7	0
gwv_96			35.80675485	2.7	1	3	4
gwv_314			31.5950271	2.7	1	6	1
gwv_866			30.04184694	2.7	2	3	3
gwv_695			28.82698438	2.7	0	5	3
hscA			26.42417352	2.7	1	4	3
gwv_485			23.42918111	2.7	1	4	3
sucA			20.3494561	2.7	1	7	0
gwv_332			20.28271617	2.7	0	3	5
acnA			19.57018513	2.7	1	5	2
infB			19.06993006	2.7	2	4	2
pheT			14.27942126	2.7	4	3	1
gwv_1352			185.0654362	2.3	0	4	3
rplX			150.1474294	2.3	0	4	3
rpsL			93.88015955	2.3	2	2	3
rpsG			86.4082733	2.3	1	3	3
rplI			80.1994963	2.3	0	2	5
gwv_873			66.73614153	2.3	1	1	5
talC			62.08394103	2.3	1	2	4
gwv_989			36.52952064	2.3	1	5	1
gwv_415			34.33779635	2.3	3	4	0
gwv_227			32.30215824	2.3	1	3	3
tig			32.28018893	2.3	1	4	2
agcS (gwv.assembly.1 618891..620235)			32.06402695	2.3	1	4	2
purB			30.36431431	2.3	2	4	1
atpD			30.17790772	2.3	1	4	2
guaB			29.97354912	2.3	0	2	5
gwv_382			29.87270914	2.3	2	5	0
gwv_976			29.05271817	2.3	1	5	1
gwv_1043			28.97535789	2.3	2	2	3
thrS			26.80496563	2.3	0	6	1
gwv_484			23.90024428	2.3	2	2	3
gwv_404			21.71444884	2.3	3	2	2
gwv_1215			14.59109858	2.3	3	1	3
yaeT			13.76564423	2.3	3	3	1
gwv_1203			8.728108414	2.3	1	3	3

gwv_113			166.4780621	2.0	1	2	3
gwv_254			113.4160354	2.0	0	4	2
gwv_953			80.67937801	2.0	1	3	2
gwv_797			61.19193632	2.0	1	2	3
gwv_308			53.09546078	2.0	2	2	2
gwv_1030			49.53555013	2.0	2	4	0
gwv_1015			43.28419185	2.0	1	3	2
gwv_109			43.05287818	2.0	1	4	1
hemH			42.43927849	2.0	0	3	3
gwv_178			38.01239925	2.0	1	3	2
gwv_580			37.95175012	2.0	0	6	0
gwv_344			37.48356021	2.0	0	1	5
gwv_283			35.45854451	2.0	0	2	4
gwv_939			32.85282152	2.0	1	3	2
secY			29.20674777	2.0	1	5	0
recA			28.95765486	2.0	2	3	1
gwv_942			28.64511776	2.0	0	4	2
gwv_688			25.70804272	2.0	2	3	1
nusA			24.56383904	2.0	0	2	4
purD			24.31350267	2.0	2	3	1
proS			23.09589783	2.0	2	2	2
gidA			20.92170359	2.0	0	3	3
gwv_888			19.61903309	2.0	1	4	1
gatA			18.83715191	2.0	2	1	3
pyrG			14.35375888	2.0	3	1	2
gwv_582			10.92713149	2.0	1	2	3
rpsO			131.7001889	1.7	0	4	1
gwv_321			99.04724952	1.7	0	4	1
rplT			89.3934139	1.7	0	2	3
rpsJ			87.43040243	1.7	1	2	2
gwv_42			61.8400838	1.7	1	3	1
gwv_1007			51.14594978	1.7	1	4	0
gwv_651			51.01612357	1.7	3	1	1
gwv_274			48.26543126	1.7	1	3	1
gwv_1028			44.36184823	1.7	0	3	2
gwv_348			43.85916772	1.7	1	1	3
gwv_722			43.85194193	1.7	0	3	2
virB9 (gwv.assembly.1 894720..895515)			43.19002583	1.7	0	3	2
gwv_90			42.49899713	1.7	0	4	1
pdxJ			39.64005534	1.7	1	2	2
gwv_561			37.51358601	1.7	1	1	3
rplB			35.97968512	1.7	1	3	1
hflC			35.85489263	1.7	1	4	0
gwv_792			35.1693724	1.7	1	2	2
gwv_1022			30.8462831	1.7	2	2	1
trxB			29.88834843	1.7	1	2	2
gwv_414			28.26750691	1.7	2	1	2
gwv_844			26.93194874	1.7	0	4	1
gwv_630			26.92838641	1.7	0	2	3
gwv_842			26.79606019	1.7	0	2	3
gwv_288			25.32158594	1.7	0	3	2
gatB			25.28421348	1.7	0	4	1

gwv_83			25.05546538	1.7	2	3	0
purF			24.77349966	1.7	0	3	2
gwv_81			24.08659211	1.7	1	1	3
gwv_191			22.98188052	1.7	1	4	0
gwv_1220			21.55524675	1.7	0	4	1
recJ			20.6633055	1.7	0	4	1
gwv_392			20.52189061	1.7	2	1	2
argD (gwv.assembly.1 846447..847623)			19.90844292	1.7	2	2	1
ispG			19.38251139	1.7	1	0	4
gwv_320			19.38152429	1.7	2	0	3
gwv_686			18.02800146	1.7	1	1	3
gwv_575			17.60571781	1.7	1	3	1
gwv_909			17.34144976	1.7	0	3	2
pyrC			16.9654557	1.7	2	2	1
pheS			16.4189823	1.7	3	1	1
gwv_653			16.259985	1.7	1	0	4
gwv_285			15.48849844	1.7	1	2	2
gwv_698			12.68960914	1.7	2	2	1
gwv_475			11.36042875	1.7	1	1	3
tkt			10.51338535	1.7	2	1	2
alaS			9.426346656	1.7	1	0	4
gwv_1065			6.423485127	1.7	2	0	3
ppdK			6.265137982	1.7	3	1	1
rpmF			137.7242785	1.3	0	1	3
rpsU			133.4409155	1.3	0	2	2
rpmI			107.0956217	1.3	1	2	1
rpsS			99.78843877	1.3	0	3	1
gwv_594			97.17979713	1.3	0	2	2
gwv_727			72.68732793	1.3	0	2	2
gwv_262			66.5071607	1.3	1	1	2
gwv_679			62.36777423	1.3	0	3	1
hscB			61.23658449	1.3	0	2	2
rplP			60.8781231	1.3	0	1	3
bfr			59.62202317	1.3	0	3	1
gwv_1319			59.20888302	1.3	0	2	2
rplS			54.3669647	1.3	1	1	2
gwv_862			48.84407551	1.3	0	1	3
gwv_955			48.56173982	1.3	0	1	3
gwv_343			44.21674204	1.3	0	1	3
ssb			43.35593387	1.3	1	1	2
gwv_1075			43.19102095	1.3	0	2	2
gwv_714			42.54805933	1.3	1	1	2
coxB			37.17608503	1.3	0	3	1
gwv_1342			35.2412702	1.3	0	3	1
gwv_637			34.78809858	1.3	0	2	2
ribB			33.09293503	1.3	1	1	2
gwv_934			30.34075579	1.3	1	0	3
gwv_824			29.9137011	1.3	1	1	2
dnaQ			29.526886	1.3	1	1	2
tsf			29.27240762	1.3	0	1	3
gwv_477			27.0925495	1.3	0	2	2
rseP			25.41528601	1.3	0	3	1

nuoD			24.68724397	1.3	0	3	1
gwv_56			23.12171142	1.3	0	3	1
gltA			22.73357718	1.3	0	3	1
gwv_1194			22.39272091	1.3	1	2	1
gwv_1332			22.03265835	1.3	2	2	0
gwv_771			21.99285089	1.3	1	2	1
gwv_1020			21.6070114	1.3	1	2	1
gwv_281			19.93161697	1.3	3	0	1
guaA			19.23082923	1.3	0	4	0
fabF			18.70037322	1.3	1	3	0
gwv_860			17.51565738	1.3	3	1	0
xerD			16.10727695	1.3	2	1	1
gwv_372			15.36175487	1.3	0	2	2
gwv_959			13.88282816	1.3	0	2	2
gwv_926			11.78564258	1.3	1	2	1
secA			10.079528	1.3	0	2	2
gwv_40			9.886742753	1.3	2	1	1
gwv_525			9.753962639	1.3	2	0	2
murE			9.242589841	1.3	2	0	2
gwv_1096			9.145280562	1.3	1	3	0
glmS			8.730303321	1.3	2	1	1
uvrB			8.332223452	1.3	2	1	1
gwv_1178			7.999888686	1.3	2	0	2
uvrA			7.350040292	1.3	1	1	2
gwv_21			6.835890749	1.3	1	2	1
gwv_136			5.813095797	1.3	0	2	2
carB			4.710483658	1.3	2	1	1
glyS			4.545312084	1.3	3	0	1
gwv_933		acrB/AcrD/AcrF family protein	2.856484571	0.7	1	1	0

Table A-4. *w*Pip proteome. *w*Mel homologs detected at the protein level in *w*Pip (Buckeye)-infected *C. pipiens* ovaries.

wMel Locus Tag	Abundance	Name	Species	Accession #	TPS > (95%)	% Cov	Most abundant peptide (95%)	Function	Unique Peptides > (95%)
WD1063	42	surface antigen Wsp	wPip	gi|190571332	128	53.5	LQYNGEVLPFK	Cell envelope biogenesis/Outer membrane	31
absent	29	Putative membrane protein	wPip	gi|190570988	68	48.8	ASQIEEVNQGVLNACVK	Cell envelope biogenesis/Outer membrane	20
WD0928	15	chaperone protein dnak (hsp70)	wPip	gi|190570602	32	38	IINEPTAAALAYGLDKK	Protein modification/degradation/Chaperones	19
WD0307	12	chaperonin groEL, 60 kDa	wPip	gi|190570503	94	62.7	EMLEDIAALTGAK	Protein modification/degradation/Chaperones	50
WD0655	10	ATP synthase F1, alpha subunit	wPip	gi|190571573	75	16	VVDALGNAIDGKGEIK	Energy production/conversion/transfer	4
WD0308	9	chaperonin groES, 10 kDa	wPip	gi|190570502	24	72.9	ESDLLAVIK	Protein modification/degradation/Chaperones	8
WD0631	8	hypothetical protein (WP0282)	wPip	gi|190570728	68	61.9	VQSVEKDAPILDFCVNK	Function unknown	29
WD1255	8	peptidoglycan-associated lipoprotein, putative	wPip	gi|190571199	20	40.9	VTLTGHTDNR	Cell envelope biogenesis/Outer membrane	8
WD0745	7	Putative outer membrane protein	wPip	gi|190571111	59	64.1	FVPYAALHYFMTDEK	Cell envelope biogenesis/Outer membrane	29
WD0683	7	translation elongation factor tu	wPip	gi|190571544	25	38.2	TTLTAAITK	Ribosome stucture/biogenesis/Translation	10
(partial)	6	ankyrin repeat domain protein	wPip	gi|190570819	18	47.8	YLIEQGANPNATDHLGR	Function unknown	11
WD0604	6	minor capsid protein E	wPip	gi|190570849	9	31.8	ALADVITDHLQLMR	Phage\Viral related proteins	4
WD0572	6	putative phage related protein	wPip	gi|190571703	12	54.1	VQEVLKDFFSPIIQKT	Phage\Viral related proteins	7
absent	5	Hypothetical protein WP0984	wPip	gi|190571376	37	65	EEVNHVNNMFGMDILNSFEGR	Function unknown	25
absent	5	Hypothetical protein WP0890	wPip	gi|190571287	23	58.8	IYNYITLAK	Function unknown	12
WD0674	5	ribosomal protein L16	wPip	gi|190571553	6	41.6	VLFEISSDVPMHLAR	Ribosome stucture/biogenesis/Translation	4
WD0016	4	translation elongation factor G	wPip	gi|190570976	14	28.1	FVPVLCGSAFK	Ribosome stucture/biogenesis/Translation	9

WD0906	3	polyribonucleotide nucleotidyltransferase	wPip	gi|190571231	16	39.8	APVAGIAMGLIK	Ribosome stucture/biogenesis/Translation	12
WD0023	3	ribosomal protein L7/L12	wPip	gi|190570969	15	93.2	EVNSTLNLK	Ribosome stucture/biogenesis/Translation	11
WD0590	3	putative phage related protein	wPip	gi|190571688	5	19.7	IVIFGPYGIGK	Phage\Viral related proteins	3
WD0879	3	thioredoxin	wPip	gi|190571104	5	60.2	AVNDQNFESEVANHK	Cellular defense mechanisms	5
WD1050	3	recA protein	wPip	gi|190571327	4	22.3	AEIEGDMGDQHMGLQAR	DNA replication/repair/packaging/Cell division	4
WD0024	2	DNA-directed RNA polymerase, beta/beta' subunits	wPip	gi|190570968	25	24	AIPGVNEENLYHLDDSGIVK	Transcription/Post-transcriptional modification	20
WD1085	2	surface antigen	wPip	gi|190571424	8	19.4	IRLDFGFPLVK	Cell envelope biogenesis/Outer membrane	7
WD0065	2	DNA-binding protein, HU family	wPip	gi|190571020	10	83.5	LKQDCVSQNIDITK	DNA replication/repair/packaging/Cell division	7
WD1271	2	enhancing lycopene biosynthesis protein 2, putative	wPip	gi|190571210	8	41.4	CFAPDINITQVMDHK	Secondary metabolite synthesis/catabolism	7
WD0658	2	DNA-directed RNA polymerase, alpha subunit	wPip	gi|190571569	7	34.9	ILQEQFQPFISSDMSYKK	Transcription/Post-transcriptional modification	7
WD0632	2	hypothetical protein (Ulp1/Peptidase_C48; WP0283)	wPip	gi|190570729	9	17	VISIDFGNPQSALDKIDGVSR	Function unknown	6
WD1090	2	ribosomal protein S1	wPip	gi|190571429	8	34.7	QIEYDPLEELIEK	Ribosome stucture/biogenesis/Translation	5
WD0631	2	hypothetical protein (WP0294)	wPip	gi|190570737	6	33.3	SAFEEDGSDDDLRR	Function unknown	5
WD0531	2	translation elongation factor Ts	wPip	gi|190571620	5	39.2	SIIEEQVK	Ribosome stucture/biogenesis/Translation	4
WD0253	2	transposase	wPip	gi|190571636	6	27.9	TTGLVDYKELETNILSSIR	DNA replication/repair/packaging/Cell division	4
WD0756	2	antioxidant, AhpC/Tsa family	wPip	gi|190570611	5	47.2	GKPAMQASDEGVADYLNSHSAEL	Cellular defense mechanisms	4
WD0790	2	Putative dnaj domain membrane protein	wPip	gi|190570961	4	18.5	DFDGLIAILK	Protein modification/degradation/Chaperones	3
WD0664	2	ribosomal protein S5	wPip	gi|190571563	4	31.8	SNDPHNIICAVFK	Ribosome stucture/biogenesis/Translation	3
WD0511	2	conserved hypothetical protein	wPip	gi|190570734	5	30	IMDEIAAFAQK	General function prediction only	3
WD0021	2	ribosomal protein L1	wPip	gi|190570971	4	37.8	FGTVTSNIAEATK	Ribosome stucture/biogenesis/Translation	3
WD0654	2	transcription elongation factor GreA	wPip	gi|190571574	4	36	DQGDLSENAEYHAAR	Ribosome stucture/biogenesis/Translation	3
WD1318	2	translation initiation factor IF-2	wPip	gi|190571749	2	13.6	ITFIDTPGHEAFTAMR	Ribosome stucture/biogenesis/Translation	1
WD0583*	2	putative phage related protein	wPip	gi|190571691	7	27.3	ILTPGGLLLLGGAPK	Phage\Viral related proteins	6
WD0585	2	putative phage related protein	wPip	gi|190571690	7	37.4	KINSIADLNGLEFTAK	Phage\Viral related proteins	5
WD0594	2	Phage related DNA methylase	wPip	gi|190571683	4	25.1	SDGTVVDGHLR	Phage\Viral related proteins	3
WD0061	2	ompA-like protein	wPip	gi|190571144	5	24.2	ILGAISYK	Cell envelope biogenesis/Outer membrane	4
WD0732	2	two component transcriptional regulator	wPip	gi|190570997	3	32.8	IGNMNINFDHR	Signal transduction	3
WD0675	2	ribosomal protein S3	wPip	gi|190571552	4	38.4	LHQDLFIR	Ribosome stucture/biogenesis/Translation	3
WD0738	2	superoxide dismutase, Fe	wPip	gi|190571001	3	20.8	LNELVENTDYQHMEIEELVTK	Cellular defense mechanisms	3
WD0589	1	putative phage related protein	wPip	gi|190571689	6	34.7	EYLNDQSSIPK	Phage\Viral related proteins	6
WD0833	1	protease DO	wPip	gi|190571439	5	21.8	INSDKDLPFVEFGNSDK	Protein modification/degradation/Chaperones	5
absent	1	Hypothetical protein WP1117	wPip	gi|190571499	6	46.3	AKTDTIPADLTAK	Function unknown	5
WD0227	1	membrane GTPase involved in stress response	wBm	gi|58584322	6	15.3	INIIDTPGHADFGGEVER	Signal transduction	4
WD0388	1	ribosomal protein S4	wPip	gi|190570680	4	40.2	IPILIEAEQKQER	Ribosome stucture/biogenesis/Translation	4
WD1319	1	N utilization substance protein A	wPip	gi|190571750	4	24.4	AITPAEVSK	Transcription/Post-transcriptional modification	4
WD0795	1	transcription termination factor Rho	wPip	gi|190570947	5	31.6	IFPAIDITK	Transcription/Post-transcriptional modification	4
WD1277	1	heat shock protein HtpG	wPip	gi|190571174	3	24.5	ELISNASDACDKLR	Protein modification/degradation/Chaperones	3
WD0678	1	ribosomal protein L2	wPip	gi|190571549	3	36.9	ATIGVVSNLDHK	Ribosome stucture/biogenesis/Translation	3
WD0320	1	trigger factor, putative	wPip	gi|190570981	3	26.6	LRFPEDYQVISLAGQEAAFSVR	Protein modification/degradation/Chaperones	3
WD0751	1	pyruvate dehydrogenase complex, E3 component, Dihydrolipoamide dehydrogenase	wPip	gi|190570560	4	29.6	DACIDAFFKK	Energy production/conversion/transfer	3
WD1210	1	succinyl-CoA synthase, beta subunit	wPip	gi|190571356	3	36.5	IVKFDIDPATGFTNLDNSK	Energy production/conversion/transfer	3
WD0762	1	peptidase, M16 family	wPip	gi|190570922	3	10.7	ELDTLLFK	General function prediction only	3
WD0174	1	ribosomal 5S rRNA E-loop binding protein Ctc/L25/TL5	wPip	gi|190571325	3	36	CSPEKIPQVIEIDLSGK	Ribosome stucture/biogenesis/Translation	3
WD0029	1	phosphoribosylamine--glycine ligase	wPip	gi|190570964	4	17.3	ANGIAAGK	Nucleotide metabolism	4
WD0532	1	ribosomal protein S2	wPip	gi|190571619	3	45.8	ILNDEDSILTKK	Ribosome stucture/biogenesis/Translation	3
WD0832	1	hflC protein	wPip	gi|190571440	4	24.1	EIRAEGEQAGQEIR	Protein modification/degradation/Chaperones	4
WD0917*	1	hypothetical protein WP0593	wPip	gi|190571002	3	29	FSDANAEGVGSPSLSK	Function unknown	3
WD0897	1	iron compound ABC transporter, periplasmic iron compound-binding protein	wPip	gi|190571080	3	29.9	KEELVHSLFDDFTK	Transporters	3
WD0391	1	ribosomal protein L28	wPip	gi|190570684	3	39.4	TFLLNLHK	Ribosome stucture/biogenesis/Translation	3
WD0681	1	ribosomal protein L3	wPip	gi|190571546	3	37.9	IGLLMTNVGHTAMYFDNSR	Ribosome stucture/biogenesis/Translation	3
WD1111*	1	hypothetical protein WP0065	wPip	gi|190570536	3	32.4	IIDETKQEIAQHIENSDVESVQLR	Function unknown	3

WD1285	1	bacterioferritin comigratory protein	wPip	gi|190571297	3	20.7	TTFLIDKK	Cellular defense mechanisms	3

Table A-5. Accession numbers. Accession numbers for WD0631/WD0632 homologs analyzed in Figs. 1c and 1e and Figs. A-1 to A-3.

	Genome	WD0631-like Accession #	WD0632-like Accession #	WD0632-like Domain	Prophage WO / WO-like Island
Type I	WOMelB	WP_010962721	WP_010962722	Ulp1/Peptidase_C48	Prophage WO
	WOSuziB	WP_044471237	WP_044471243	Ulp1/Peptidase_C48	Prophage WO
	WORiB	WP_012673191	CP001391*	Ulp1/Peptidase_C48	Prophage WO
	WOHa1	WP_015588933	WP_015588932	Ulp1/Peptidase_C48	Prophage WO
	WOSol	AGK87106	AGK87078	Ulp1/Peptidase_C48	Prophage WO
	WORecB	WP_038198916	JQAM01000018*	Ulp1/Peptidase_C48	Prophage WO
	WOPip1	WP_012481787	WP_012481788	Ulp1/Peptidase_C48	Prophage WO
	WOVitA4	PRJNA213627*	PRJNA213627*	Ulp1/Peptidase_C48	Prophage WO
Type II	WORiC	WP_012673228	WP_012673227	None	Prophage WO
	WOSuziC	WP_044471252	WP_044471251	None	Prophage WO
Type III	*w* No	WP_015587806	WP_015587805	Nuclease	WO-like Island
	w AlbB	WP_006014162	WP_006014164	Nuclease	WO-like Island
	w VitA	PRJNA213627*	PRJNA213627*	Nuclease	WO-like Island

**Wolbachia*'s CI gene types are classified according to their WD0632-like functional domain. Type I features the Peptidase_C48 domain, Type II lacks an annotated C-terminal domain, and Type III encodes the PD-(D/E)XK nuclease domain. All genes are either associated with the prophage WO genome or located within WO-like islands on the *Wolbachia* chromosome. (*) indicates that the accession number refers to a nucleotide sequence; proteins were translated using Geneious.

Table A-6. Primers. Primers utilized in this study are listed. F = forward primer, R = reverse primer.

Primer	Sequence	Product Length
Rp49_F	CGGTTACGGATCGAACAAGC	154bp

Rp49_R	CTTGCGCTTCTTGGAGGAGA	
groELstd_F	GGTGAGCAGTTGCAAGAAGC	923bp
groELstd_R	AGATCTTCCATCTTGATTCC	
groEL_F	CTAAAGTGCTTAATGCTTCACCTTC	97bp
groEL_R	CAACCTTTACTTCCTATTCTTG	
WD0034_F	GGAAGAAACTTGCACACCACTTAC	151bp
WD0034_R	TGCTCTCCGACCATCTGGATATTT	
WD0508_F	TAGAGATCTAGCTTGCGGACAAGA	204bp
WD0508_R	TCCTTAACTAAACCCTTTGCCACC	
WD0625_F	GAGCCATCAGAAGAAGATCAAGCA	120bp
WD0625_R	TTCTCGAAAGCTGAAATAGCCTCC	
WD0631_F	TGTGGTAGGGAAGGAAAGAGGAAA	111bp
WD0631_R	ATTCCAAGGACCATCACCTACAGA	
WD0632_F	TGCGAGAGATTAGAGGGCAAAATC	197bp
WD0632_R	CCTAAGAAGGCTAATCTCAGACGC	
WD0640_F	CTACAACCTCATCGAAGCGAATCT	144bp
WD0640_R	CTGCAGAAGCTTTGGAAAAATGGG	
WD0508opt_F	GACGTGCTGATCAAGAGCCT	136bp
WD0508opt_R	TGCCCACTGTCTTCAGGATG	
WD0625opt_F	CGCGAGATGGATGACCTGAA	180bp
WD0625opt_R	CTCGCGCTCACTATGTCCAA	
WD0631opt_F	GGTGGATAGTCAGGGCAACC	191bp
WD0631opt_R	AAAAGTACTCCACGCCCTCG	
WD0632opt_F	CCTGCCCTACATTACACGCA	159bp
WD0632opt_R	GGCGACAGATCCAGGTCAAT	
Wolb_F	GAAGATAATGACGGTACTCAC	990bp
Wolb_R3	GTCACTGATCCCACTTTAAATAAC	

Table A-7. Exact p-values. The exact p-values for all statistical calculations, along with method used, are listed.

Figure	p-value	Comparison	Test
Fig. 2	0.0022	WD0631 1 day old vs 7 days old	Mann-Whitney (two-tailed)
	0.0022	WD0632 1 day old vs 7 days old	Mann-Whitney (two-tailed)
	0.0022	WD0508 1 day old vs 7 days old	Mann-Whitney (two-tailed)
	0.0411	WD0625 1 day old vs 7 days old	Mann-Whitney (two-tailed)
	0.0022	WD0640 1 day old vs 7 days old	Mann-Whitney (two-tailed)
Fig. 3a	0.0022	(-) M x (-) F vs (+) M x (-) F	ANOVA (Kruskal-Wallis with Dunn's correction)
Fig. 3b	<0.0001	(-) M x (-) F vs (+) M x (-) F	ANOVA (Kruskal-Wallis with Dunn's correction)

	<0.0001	(-) M x (-) F vs (-)WD0631,WD0632 M x (-) F	ANOVA (Kruskal-Wallis with Dunn's correction)
	<0.0001	(+) M x (-) F vs (-)WD0631,WD0632 M x (+) F	ANOVA (Kruskal-Wallis with Dunn's correction)
	<0.0001	(-)WD0631,WD0632 M x (-) F vs (-)WD0631,WD0632 M x (+) F	ANOVA (Kruskal-Wallis with Dunn's correction)
Fig. 3c	0.0390	(+) M x (-) F vs (+)WD0631 M x (-) F	ANOVA (Kruskal-Wallis with Dunn's correction)
	0.0047	(+) M x (-) F vs (+)WD0632 M x (-) F	ANOVA (Kruskal-Wallis with Dunn's correction)
	0.0102	(+)WD0631 M x (-) F vs (+)WD0631,WD0632 M x (-) F	ANOVA (Kruskal-Wallis with Dunn's correction)
	<0.0001	(+) M x (-) F vs (+)WD0631,WD0632 M x (-) F	ANOVA (Kruskal-Wallis with Dunn's correction)
Fig. 4g	<0.0001	(-) M x (-) F vs (+) M x (-) F	Fisher's exact
	<0.0001	(-) M x (-) F vs (-)WD0631,WD0632 M x (-) F	Fisher's exact
	0.0002	(+) M x (-) F vs (-)WD0631,WD0632 M x (-) F	Fisher's exact
Fig. 4h	<0.0001	(-) M x (-) F vs (+) M x (-) F	Fisher's exact
Ext. Data Fig. 4a	0.0002	Testes (1 day old) vs Testes (7 days old)	ANOVA (Kruskal-Wallis with Dunn's correction)
Ext. Data Fig. 4b	0.0007	Testes (4 days old) vs Testes (7 days old)	ANOVA (Kruskal-Wallis with Dunn's correction)
Ext. Data Fig. 4d	0.0104	Older brothers vs younger brothers	Mann-Whitney (two-tailed)
Ext. Data Fig. 4e	<0.0001	(+) M x (+) F vs (+) M x (-) F (older)	ANOVA (Kruskal-Wallis with Dunn's correction)
	<0.0001	(+) M x (+) F vs (+) M x (-) F (younger)	ANOVA (Kruskal-Wallis with Dunn's correction)
Ext. Data Fig. 5a	0.0004	(-) M x (-) F vs (+) M x (-) F	ANOVA (Kruskal-Wallis with Dunn's correction)
	0.0220	(-) M x (-) F vs (+) M x (+) F	ANOVA (Kruskal-Wallis with Dunn's correction)
Ext. Data Fig. 8a	<0.0001	(-) M x (-) F vs (+) M x (-) F	ANOVA (Kruskal-Wallis with Dunn's correction)
Ext. Data Fig. 8b	0.0032	(+) M x (-) F vs (+)WD0625,WD0632 M x (-) F	ANOVA (Kruskal-Wallis with Dunn's correction)
	0.0002	(+) M x (-) F vs (+)WD0632 M x (-) F	ANOVA (Kruskal-Wallis with Dunn's correction)
	0.0011	(+)WD0625 M x (-) F vs (+)WD0625,WD0632 M x (-) F	ANOVA (Kruskal-Wallis with Dunn's correction)
	<0.0001	(+)WD0625 M x (-) F vs (+)WD0632 M x (-) F	ANOVA (Kruskal-Wallis with Dunn's correction)
Ext. Data Fig. 8c	<0.0001	(-) M x (-) F vs (+) M x (-) F	ANOVA (Kruskal-Wallis with Dunn's correction)
	0.0023	(+) M x (-) F vs (-)WD0625,WD0632 M x (-) F	ANOVA (Kruskal-Wallis with Dunn's correction)
Ext. Data Fig. 10a	0.0002	WT vs WD0508	Mann-Whitney (two-tailed)
Ext. Data Fig. 10b	0.0003	WT vs WD0632	ANOVA (Kruskal-Wallis with Dunn's correction)
	0.0004	WD0625 vs WD0632	ANOVA (Kruskal-Wallis with Dunn's correction)

Ext. Data Fig. 10c	<0.0001	WT vs WD0632	ANOVA (Kruskal-Wallis with Dunn's correction)
	0.0334	WT vs WD0631,WD0632	ANOVA (Kruskal-Wallis with Dunn's correction)

* M = male, F = female, + = *Wolbachia* infected, - = *Wolbachia* uninfected

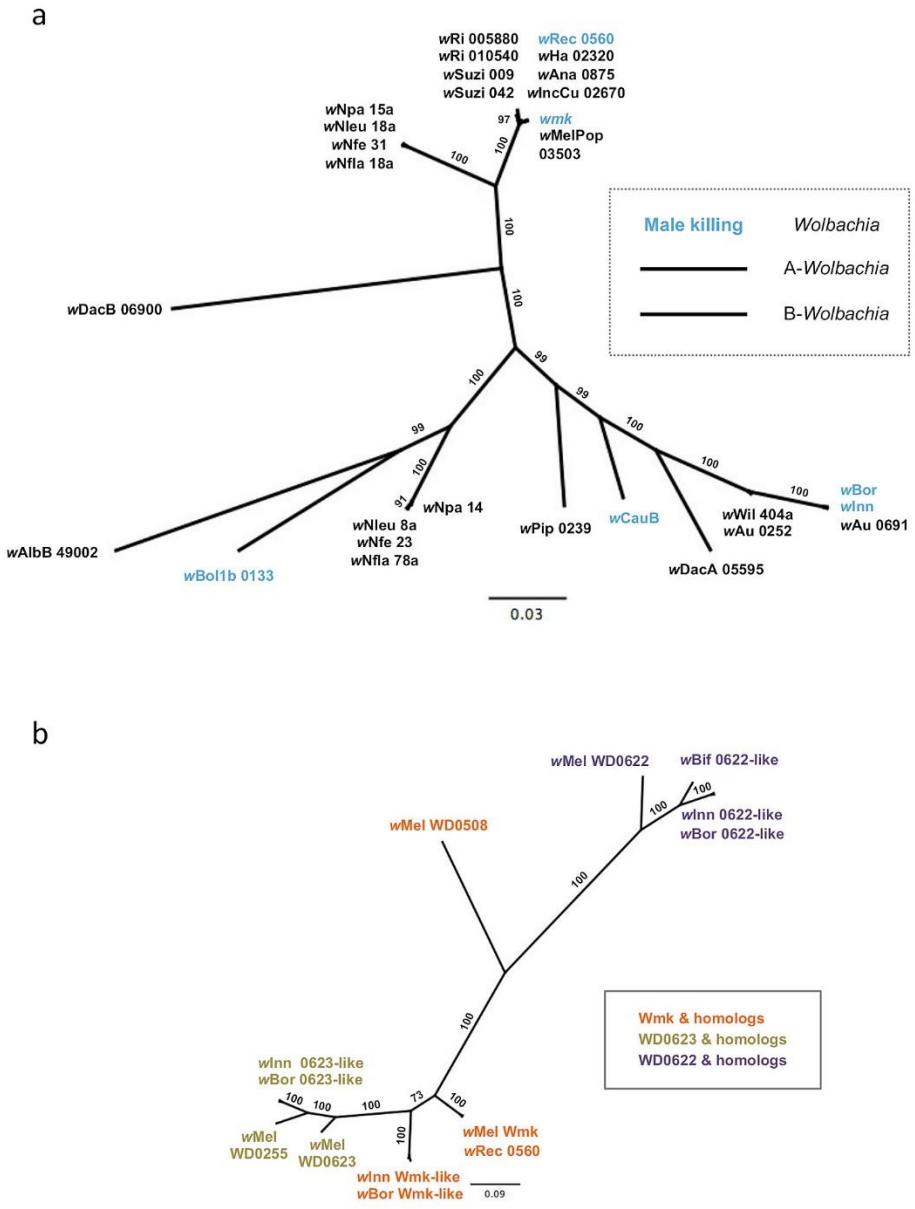


Figure B-1. Comparative genomics of the *wmk* gene and protein.
(a) Phylogeny of full-length *wmk* gene, based on an 893-bp alignment and GTR+I+G model of evolution. Full-length *wmk* homologs are widespread throughout prophage WO-containing *Wolbachia* strains, some of which are male-killing strains. Like many WO-associated genes, including CI factors *cifA* and *cifB*, the *wmk* phylogeny does not support evolution with the *Wolbachia* chromosome because homologs in A- and B-*Wolbachia* do not cluster according to supergroup. *Wolbachia* supergroups are illustrated as either black (A-*Wolbachia*) or red (B-*Wolbachia*) branches. *wmk* (WD0626) and homologs from male-killing strains are highlighted in cyan. Consensus support values are shown on the branches. The tree includes all taxa that are reciprocal best hits of *w*Mel *wmk*. See S5 Table for accession numbers and BLASTn E-values. (b) A Bayesian phylogeny of Wmk protein and homologous peptides from *w*Mel and sequenced male-

killing strains in *Drosophila*, based on a 168 aa-alignment using the JTT+G model of evolution. It shows that homologs in these taxa cluster according to gene synteny within prophage WO genomes (see Fig. 4-1). Consensus support values are shown on the branches. Colors correspond to Fig. 4-1. Accession numbers and BLASTP E-values are provided in Table B-4.

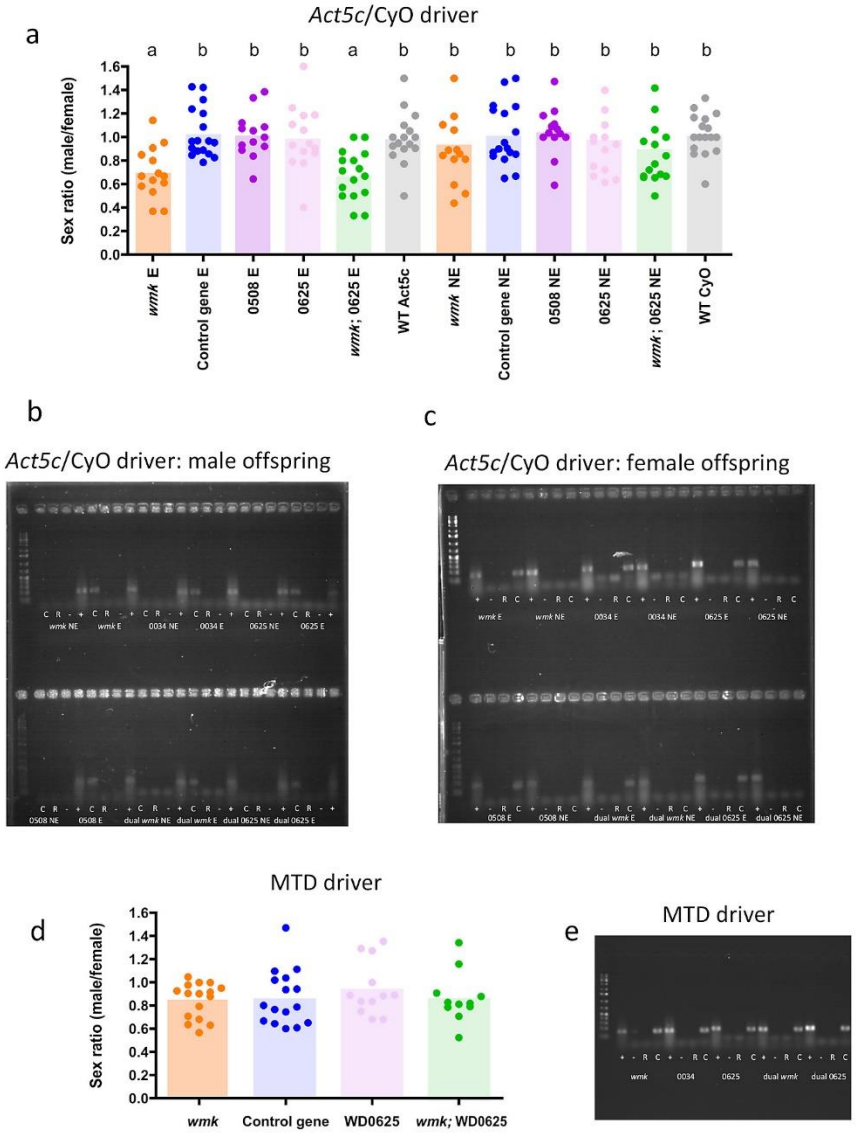


Figure B-2. *Act5c* and MTD expression of transgenes other than *wmk* do not cause a sex ratio bias.

(a) Sex ratios were quantified for *wmk*, the control gene (WD0034), WD0625, WD0508, or dual *wmk*;WD0625 transgenes expressed with the *Act5c*/CyO driver. Expressing and non-expressing flies of each genotype are siblings. Each point represents the offspring of one vial of 10 mothers and 2 fathers. A biased sex ratio only results when *wmk* is expressed. Average N per vial is 78. Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction with only the non-expressing flies or only the expressing flies. Groups labeled "a" are significantly different compared to groups labeled "b". Non-expressing flies are non-significant. Bars represent the average sex ratio. E = expressing, NE = non-expressing, *Act5c* = *Act5c* gene present, CyO =

CyO chromosome present. This experiment has been performed once. (b) Transgenes are expressed in *Act5c* (E) adult males but not their CyO (NE) brothers as demonstrated by cDNA generated from males. Samples were taken from offspring of parental siblings from the experiment in (a). Samples were from pooled, whole-body, adult extractions of three individuals from each genotype. (c) Transgenes are expressed in *Act5c* (E) females, but not their CyO (NE) sisters as demonstrated by cDNA generated from females. See (b) for details. Both (b) and (c) have been performed once. (d) Sex ratios were similarly quantified for the listed transgenes using a maternal triple driver (MTD) where expression was driven in the mother throughout oogenesis and offspring were laid with the expressed products loaded into the eggs. Each point represents the offspring of a vial of 10 mothers and 2 fathers. Average N per vial is 74. Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction. Bars represent the average sex ratio. This experiment has been done once. (e) Transgenes are expressed in all adult offspring from the MTD driver as demonstrated by cDNA generated from siblings of mothers from the experiment in (d). Samples are from pooled, dissected ovaries of six flies. This experiment has been performed once. The meanings of notations in the gels are as follows: "dual *wmk*" indicates *wmk*;0625 co-expressing flies measured with *wmk* primers, "dual 0625" indicates *wmk*;0625 co-expressing flies measured with 0625 primers, "C" indicates cDNA, "R" indicates RNA, "-" indicates negative control, "+" indicates positive control.

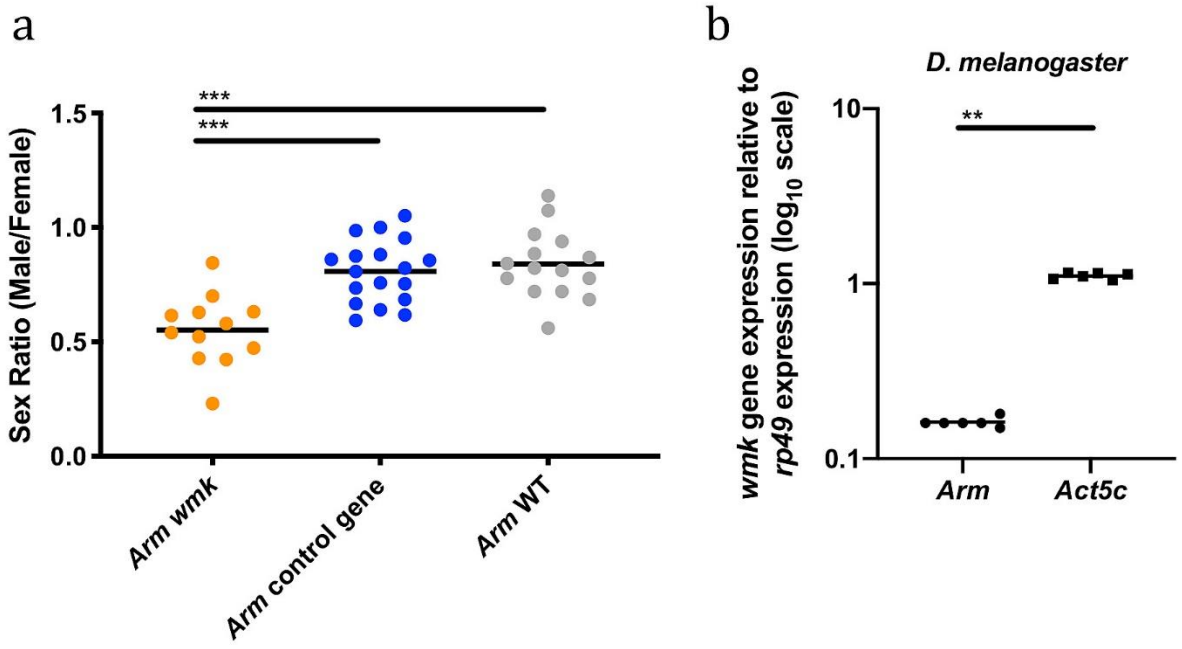


Figure B-3. The *wmk* phenotype can be induced with both the *arm* and *Act5c* drivers despite differences in expression levels.

(a) Sex ratios were quantified for *wmk*, the control gene (WD0034), and WT flies. The *armadillo* (*arm*) driver is homozygous, so all offspring express the gene. Each point represents the offspring of one vial of 10 mothers and 2 fathers. A biased sex ratio results only when *wmk* is expressed, as with *Act5c* (Fig. 4-2). Average N per vial is 73. Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction. This experiment has been performed twice. (b) Graph of transgenic *wmk* expression compared to *Drosophila* housekeeping gene *rp49*. Each point (n = 6) represents a pool of 30 embryos from a set of 10 mothers and 2 fathers. Values denote $2^{-\Delta Ct}$.

Statistics are based on a Mann-Whitney U test. This experiment has been performed once. **p<0.01, ***p<0.001.

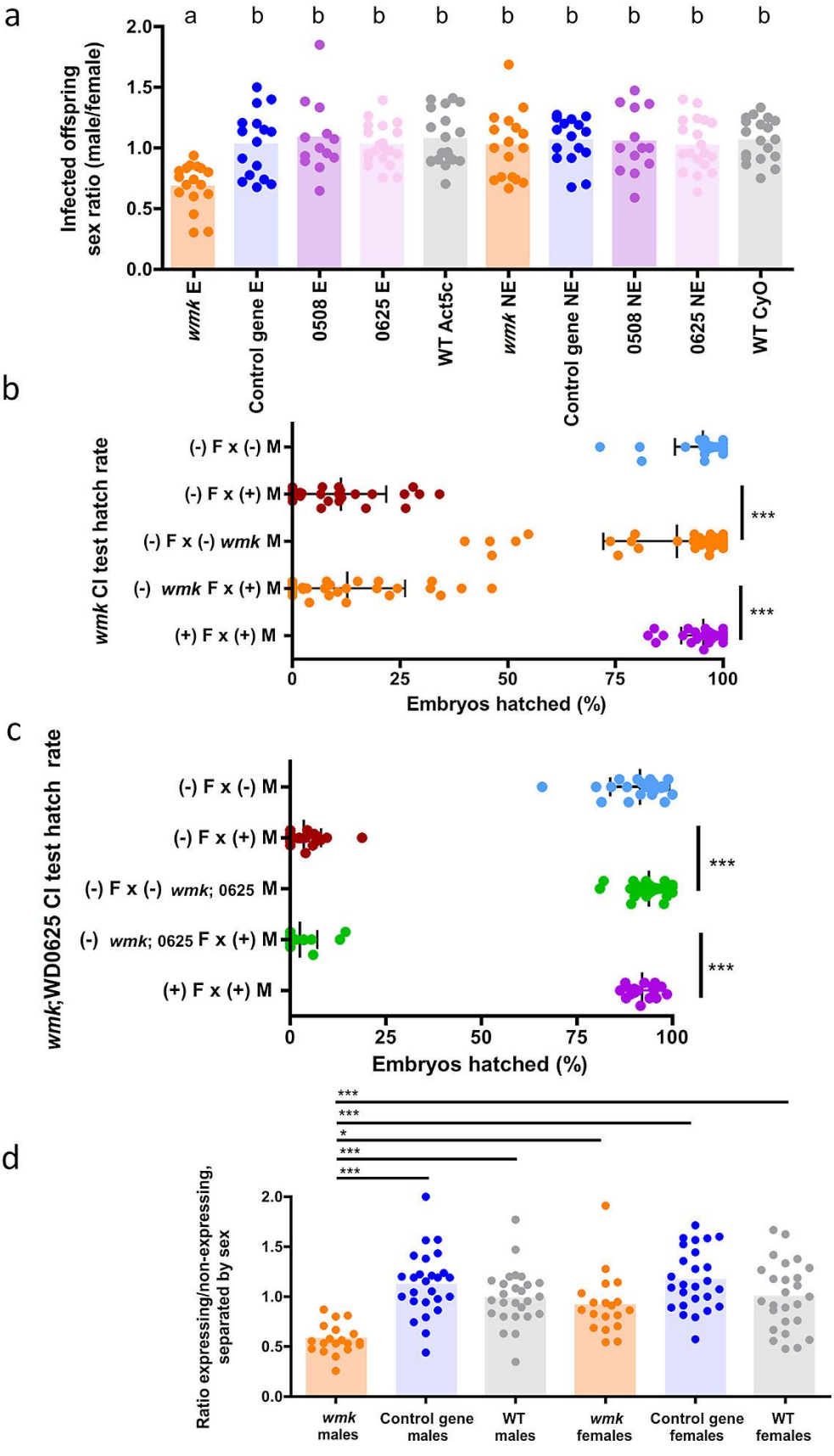

a
b
c
d
Infected offspring sex ratio (male/female)
a b b b b b b b b b
wmk E
Control gene E
0508 E
0625 E
WT Act5c
wmk NE
Control gene NE
0508 NE
0625 NE
WT CyO
wmk CI test hatch rate
(-) F x (-) M
(-) F x (+) M
(-) F x (-) wmk M
(-) wmk F x (+) M
(+) F x (+) M
***
Embryos hatched (%)
wmk;WD0625 CI test hatch rate
(-) F x (-) M
(-) F x (+) M
(-) F x (-) wmk; 0625 M
(-) wmk; 0625 F x (+) M
(+) F x (+) M
***
Embryos hatched (%)
Ratio expressing/non-expressing, separated by sex
***
***
*
***
***
wmk males
Control gene males
WT males
wmk females
Control gene females
WT females

Figure B-4. The *wmk* phenotype is not due to other forms of reproductive parasitism and it does not induce or rescue CI.

(a) Resulting offspring sex ratios from infected mothers are shown here. Sex ratios of infected offspring of the indicated genotypes demonstrate that an infected background does not rescue or alter the *Act5c* driver-induced phenotype, which would be a characteristic of CI. Each point represents the offspring of a single vial of mothers and fathers. Average N per vial is 105 offspring. The group labeled "a" is significantly different compared to groups labeled "b". Bars represent the average sex ratio. E = expressing, NE = non-expressing, *Act5c* = has *Act5c* gene, CyO = has CyO chromosome. (b) Hatch rate of offspring with parents expressing *wmk* under a *nanos* driver (expression in the gonads) in either fathers or mothers to test CI induction or rescue, respectively. Expression in males does not recapitulate wild type (WT) CI, and expression in females does not recapitulate rescue. Each dot represents the hatch rate of offspring of a single male and female, N = 24–36 crosses per group. Bars indicate average ± s. d. (c) Same as (b), but offspring have parents dually expressing *wmk*; WD0625 under a *nanos* driver (expression in gonads) in either fathers or mothers to test CI induction or rescue, respectively. Dual expression in males does not recapitulate WT CI, and dual expression in females does not recapitulate rescue CI. (d) Ratio of expressing to non-expressing flies (same flies as Fig. 4-2) broken down by sex (ie, expressing males compared to non-expressing males, expressing females compared to non-expressing females). Each dot represents a comparison of sibling (brothers or sisters) offspring from a single vial of mothers and fathers. Bars represent the average ratio. The *wmk* male ratio is reduced, but *wmk* females are not significantly increased compared to controls. This indicates a loss of *wmk*-expressing males without a corresponding increase in females, suggesting male killing rather than feminization. Statistics for (a) and (d) experiments are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction. Statistics for (b) and (c) were performed with a Mann-Whitney U test between each of the two groups indicated by the significance bars. These experiments have all been performed once. *$p<0.05$, **$p<0.01$, ***$p<0.001$. (-), no *Wolbachia* infection; (+), *Wolbachia* infection; blue, normal; red, CI cross; purple, rescue cross; orange, *wmk* cross; green, dual *wmk*; WD0625.

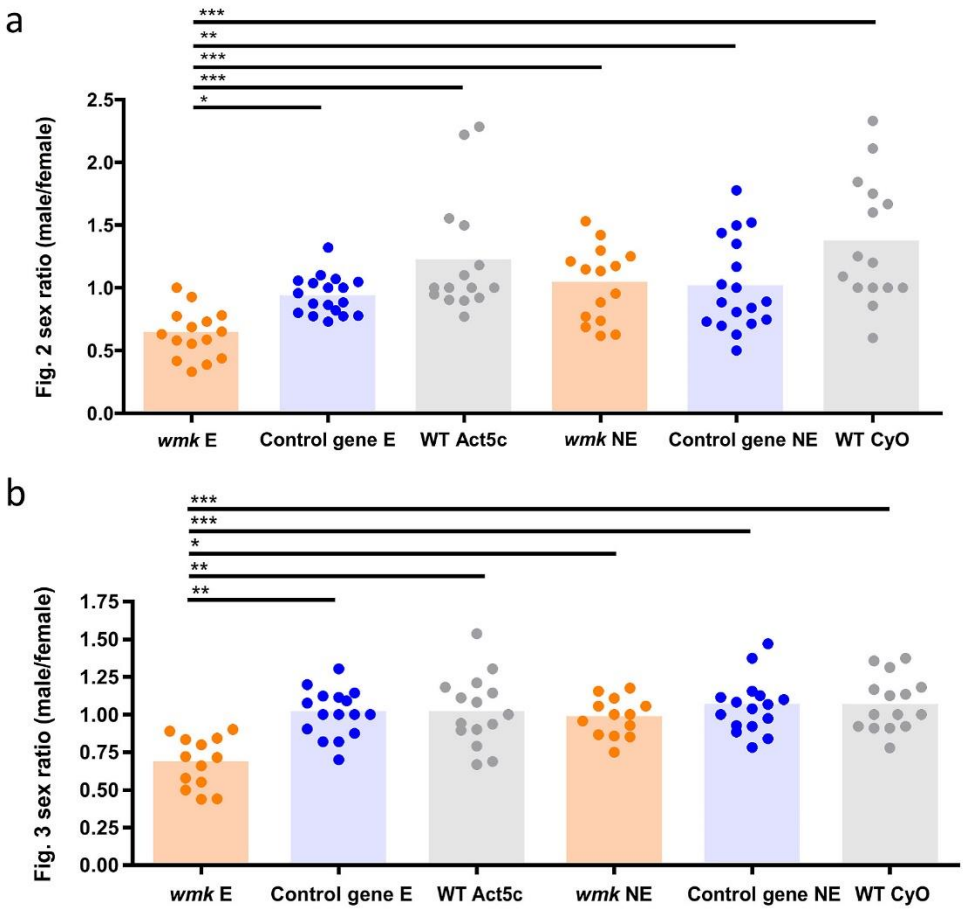


Figure B-5. The corresponding sex ratios of all experiments are female-biased.
Alongside the experiments in Figs. 4-3 and 4-4, sex ratios were measured. (a) Graph of the adult offspring sex ratio from the cytology experiment in Fig. 4-3. Each point represents the offspring of a single vial of mothers and fathers. This was measured with offspring of siblings to the flies used to lay eggs in Fig. 4-3. Average N is 79 adult offspring per cross of 10 mothers and 2 fathers. Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction. This experiment has been done once. (b) Graph of the adult sex ratio from the experiment in Fig. 4-4. Each point represents the offspring of a single vial of mothers and fathers. This was measured with offspring of siblings to the flies used to lay eggs in Fig. 4-4. Average N is 85 adult offspring per vial. Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction. E = expressing, NE = non-expressing, *Act5c* = has *Act5c* gene, CyO = has CyO chromosome. *p<0.05, **p<0.01, ***p<0.001. This experiment has been done once. All bars represent the average sex ratio.

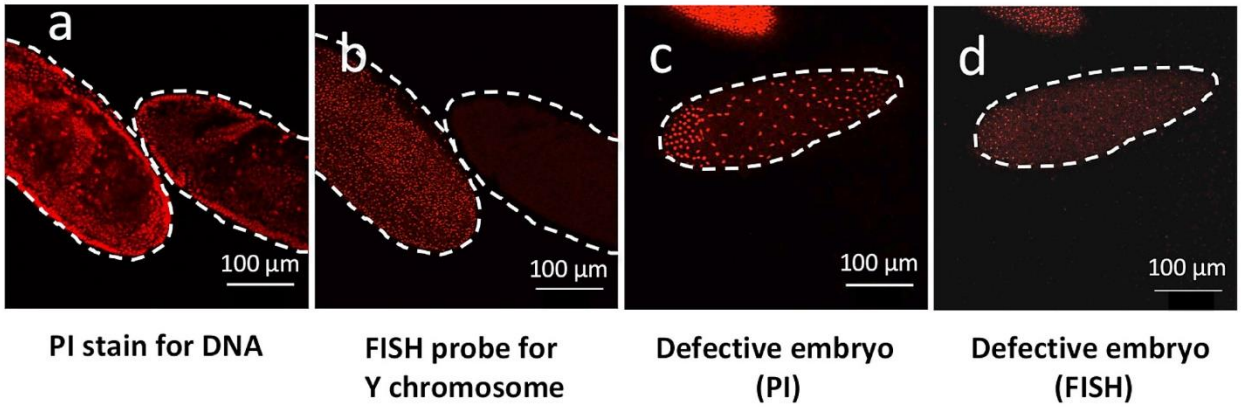


Figure B-6. Representative images of FISH staining of Y chromosome from data in Fig. 4-3.

These images were taken as a part of the experiment described in Fig. 4-4. (a) Image of two normal control gene embryos approximately 4 h after egg deposition (AED) stained for DNA with PI. (b) Image of the same embryos as (a) stained with a Cy5-conjugated FISH probe specific to the Y chromosome. The left embryo is male, the right embryo is female. (c) Image of a *wmk* embryo 3–4 h AED stained with PI showing local mitotic failure and chromatin bridging. (d) Image of the same embryo as (c) stained with the Y chromosome probe, showing it is male.

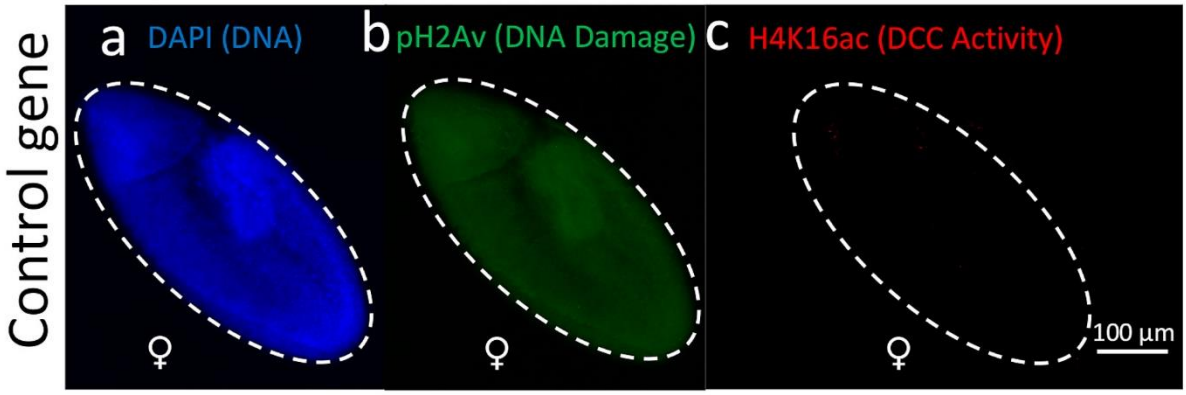


Figure B-7. Control gene females do not show DNA damage.

These images were taken as a part of the experiment described in Fig. 4-5, and all three are of the same embryo. (a) Image of a normal control gene female stained with DAPI for DNA at 4–5 h AED. (b) Image of the embryo stained with an antibody for pH2Av, demonstrating only background signal. (c) Image of the embryo stained with an antibody for H4K16ac, demonstrating no detectable signal.

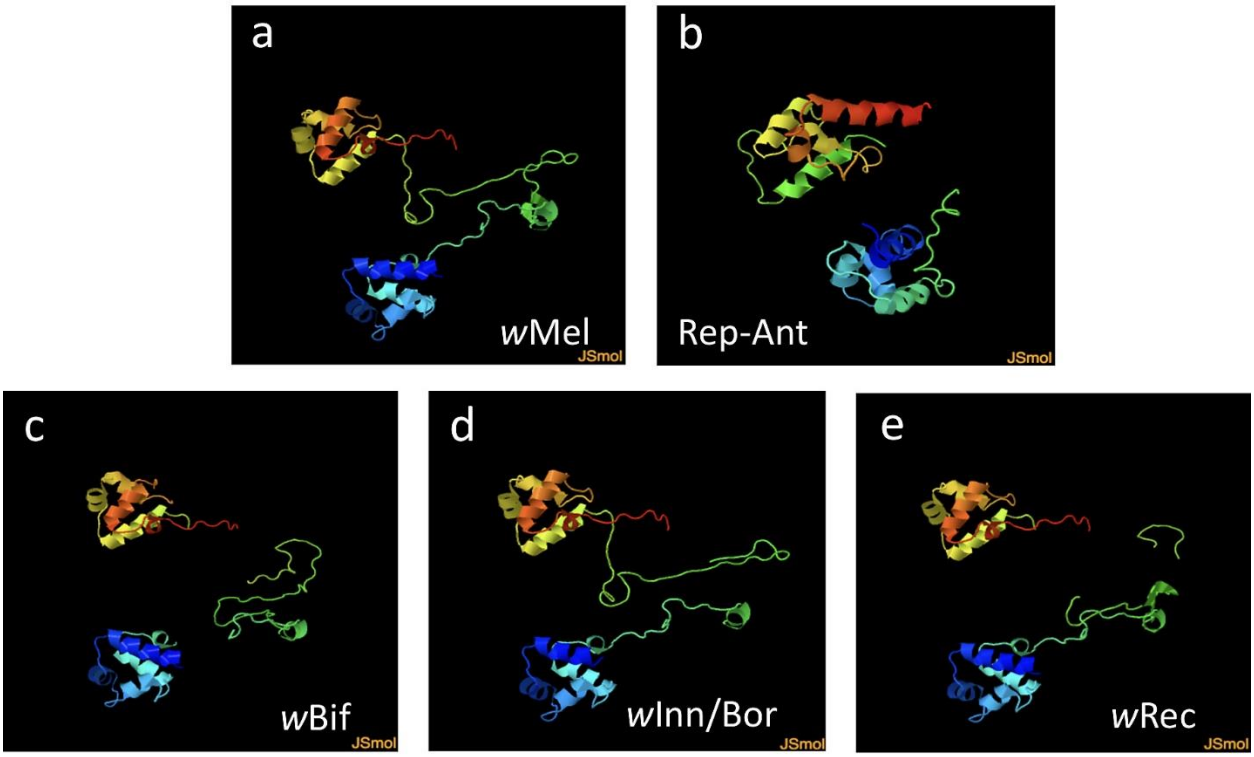


Figure B-8. Predicted protein architecture of Wmk homologs and homology to a phage repressor.

(a) Image of the most likely 3D structure of Wmk from *w*Mel determined by the Phyre2 web portal. 66% of residues modeled at 99.9% confidence. (b) The known protein structure with the most shared similarity across all homologs, with the highest sequence identity and confidence, is a known phage DNA-binding transcriptional repressor. Namely, it is the Rep-Ant complex from *Salmonella*-temperate phage, modeled here (99.8% confidence and 19% residue identity compared to *w*Mel Wmk). Other top results were also almost exclusively transcriptional regulators and DNA-binding proteins. The Rep-Ant complex is comprised of two separate, dimerized peptides, and does not include the linker region of Wmk shown in green in (a). (c) Image of the most likely 3D structure of Wmk from *w*Bif determined by Phyre2. 73% of residues modeled at 99.9% confidence. (d) Image of the most likely 3D structure of Wmk from *w*Inn/*w*Bor (same sequence) determined by Phyre2. 59% of residues modeled at 99.9% confidence. (e) Image of the most likely 3D structure of Wmk from *w*Rec determined by Phyre2. 62% of residues modeled at 99.9% confidence. All images are colored in order of the rainbow from N terminus (red) to C terminus (blue). Although there are no breaks in input sequence, some breaks are shown in the models because of low confidence in modeling in those regions (the linker region between the two alpha helix bundles).

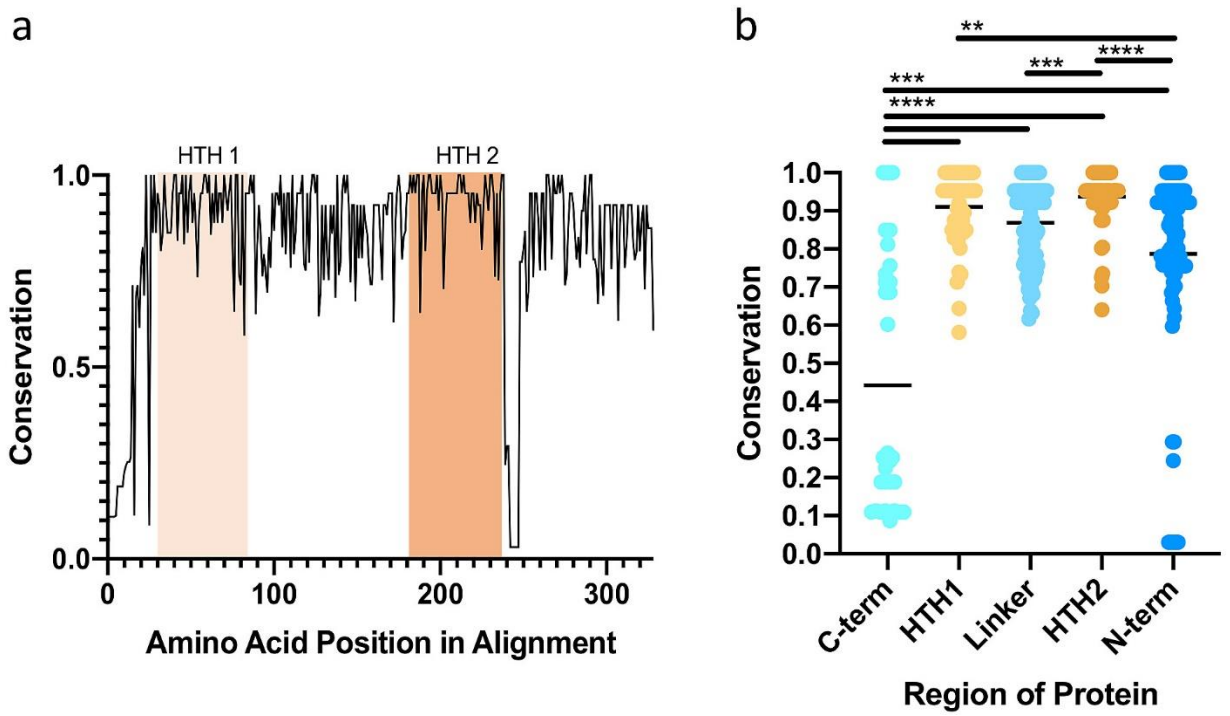


Figure B-9. Wmk amino acid identity is more conserved in the DNA-binding domains than certain other regions of the protein.
(a) Level of amino acid conservation is shown across the length of the amino acid alignment of 31 Wmk homologs. The homologs used in the analysis include all those shown in Fig. B-1 with the addition of the *w*Bif homolog. A score of 1 indicates complete conservation across homologs while a score of 0 indicates all homologs have different amino acid identities in that location. The two HTH DNA-binding domains are highlighted in shades of orange for reference. (b) Amino acid conservation from the same set of data as (a) is shown in a different format here, where each dot represents the conservation of a particular amino acid position within a designated region of the protein. Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction. Bars indicate mean values. Colors of HTH regions correspond to (a) and shades of blue are used to distinguish other regions. **p<0.01, ***p<0.001, ****p<0.0001.

Table B-1. Full details of comparative genomics analysis for male-killing gene candidates.
All *w*Bif phage genes are listed with scaffold numbers and annotations. Mobile elements were removed from further analysis, but all other genes were evaluated for presence in other strains, disruptions in other strains, E^{-20} thresholds. Remaining genes that fit all criteria were included in the final candidate list (Table 4-1).

Contig	All *w*Bif phage genes	Blast	Results					
		1:1 *w*Mel Blastp						
***w*Bif Query**	***w*Bif Annotation**	**Ref-Seq Coverage**	**E-value**	**Pairwise % Identity**	***w*Mel locus tag**	***w*Rec locus tag(s)**	***w*Inn**	***w*Bor**
Final Gene	**Candidates**							
4.4	Ankyrin repeat	93%	3.E-35	32%	WD0550	wrec0541	Scaffold 84	Scaffold 6

4.6	Transcriptional regulator	91%	9.E-25	28%	WD0626	wrec0560	Scaffold 13	Scaffold 6
4.8	Rpn	100%	0.E+00	85%	WD0627	wrec0561	Scaffold 100	Scaffold 10
4.9	HP	97%	1.E-98	81%	WD0628	wrec0562	Scaffold 19	Contig 31
4.10	CifA	99%	0.E+00	70%	WD0631	wrec0566	Scaffold 19	Contig 31
12.1	Rpn	100	0.E+00	98%	WD0296	wrec0561	Scaffold 100	Scaffold 10
12.31	PLDc-Nuc	91%	4.E-64	55%	WD1243	wrec1232	Contig 3	Contig 31
REMOVED: E-20 threshold								
12.11	HP with SMC_N	9%%	2.E-19	62%	WD0234	wrec0234	Scaffold 86	Contig 31
12.21	HP with SMC_N	64%	6.E-18	33%	WD0462	wrec0457	Scaffold 30/Scaffold 5	Contig 23
REMOVED: Disrupted in another strain		***w*Mel locus**	***w*Rec**	***w*Inn**	***w*Bor**	***w*Bif**		
4.5	Ankyrin repeat	WD0147	DISRUPTED			PARTIAL		
4.13	Ankyrin repeat	WD0073	wrec0060	DISRUPTED		PARTIAL		
12.6	Ankyrin repeat*	WD0385	DISRUPTED			PARTIAL		
12.10	Ankyrin repeat*	WD0766	DISRUPTED			PARTIAL		
11.6	HP, partial	WD1278	wrec1268	DISRUPTED		DISRUPTED		
11.8	HP, partial	WD1278	wrec1268	DISRUPTED		DISRUPTED		
12.2	HP, partial	WD0295	wrec0284	Scaffold 100	Scaffold 10	DISRUPTED		
12.3	HP, partial	WD0295	wrec0284	Scaffold 100	Scaffold 10	DISRUPTED		
REMOVED: Mobile element								
1.1	IS5, orfB							
4.1	Reverse transcriptase, group II intron origin							
4.11	IS5, orfA							
4.12	IS5, orfB							
11.1	IS5, orfA							
11.2	IS5, orfB							
11.3	Reverse transcriptase, group II intron origin							
11.4	Reverse transcriptase, group II intron origin							
11.9	IS5, orfA							
11.10	IS5, orfB							
12.4	Putative IS4							

12.7	IS256							
12.8	IS256							
12.9	putative IS5							
12.12	IS481							
12.13	IS481							
12.14	IS481							
12.15	Reverse transcriptase, group II intron origin							
12.16	Reverse transcriptase, group II intron origin							
12.19	IS5, orfB							
12.20	IS5, orfA							
12.23	Recombinase, partial							
12.24	Recombinase, partial							
12.25	IS5, orfA							
12.26	IS5, orfB							
12.32	Reverse transcriptase, group II intron origin							
REMOVED: Not present in all query strains		***w*Mel locus**	***w*Mel**	***w*Rec**	***w*Inn**	***w*Bor**		
1.2	membrane protein	WD0621		ABSENT				
1.3	UDP-glucose 6-dehydrogenase	WD0620		ABSENT				
1.4	GlpT/PgpT/UhpT transporter family protein	WD0619		ABSENT				
1.5	L-allo-threonine aldolase	WD0618		ABSENT				
1.6	L-allo-threonine aldolase	WD0617		ABSENT				
1.7	ABC transporter	WD0616		ABSENT				
1.8	HP	WD0615		ABSENT				
1.9	HP	WD0614		ABSENT				
1.10	glycosyl transferase	WD0613		ABSENT				
1.11	NAD-dependent epimerase/dehydratase family protein	WD0612		ABSENT				
1.12	UDP-N-acetylglucosamine pyrophosphorylase-related protein CDS	WD0611		ABSENT				
4.2	Patatin, partial	WD0565		ABSENT				
4.3	HP		ABSENT	ABSENT				
4.7	DNA ligase		ABSENT	ABSENT				
11.5	Patatin, disrupted			ABSENT				

11.7	Lipase		ABSENT	ABSENT				
11.11	SNF2 Helicase	WD0610		ABSENT				
11.12	DEAD-like helicase		ABSENT	ABSENT				
12.5	HP		ABSENT	ABSENT				
12.17	HP		ABSENT	ABSENT				
12.18	HP		ABSENT	ABSENT				
12.22	Ankyrin repeat	WD0633		ABSENT				
12.27	Transcriptional regulator, partial	WD0622		ABSENT				
12.28	Transcriptional regulator, partial	WD0622		ABSENT				
12.29	HP		ABSENT	ABSENT				
12.30	Patatin	WD0565		ABSENT				
13.1	Recombinase	WD0519		wrec0511	ABSENT	ABSENT		
13.2	HP		ABSENT	ABSENT				
13.3	SecA		ABSENT	ABSENT				
13.4	SecA		ABSENT	ABSENT				

Table B-2. Homologs of Wmk from related bacterial strains.
All non-*w*Mel *Wolbachia* Wmk homologs with a reciprocal BLASTp E-value of E^{-15} or above were included, and all have a reciprocal best BLAST (RBB) of the Wmk protein in *w*Mel. Accession numbers for NCBI are included.

Organism	Accession	BLASTp E-value	RBB in *w*Mel	Reciprocal BLASTp E-value to Wmk
Wolbachia endosymbiont of *Wuchereria bancrofti* (*w*Wb)	WP_088415462.1	1.00E-36	WD0626	6.00E-42
Wolbachia endosymbiont strain TRS of *Brugia malayi* (*w*Bm)	AAW70776.1	5.00E-22	WD0626	2.00E-27
Wolbachia endosymbiont of *Brugia malayi* (*w*Bm)	WP_050707658.1	2.00E-19	WD0626	5.00E-25
Wolbachia endosymbiont of *Pratylenchus penetrans* (*w*Ppe)	WP_070064999.1	1.00E-13	WD0626	4.00E-19
Rickettsiales bacterium Ac37b	WP_038602295.1	5.00E-11	WD0626	2.00E-16
Ehrlichia canis str. Oklahoma	AAK28679.1	8.00E-11	WD0626	6.00E-16
Ehrlichia canis str. Jake	WP_011305001.1	9.00E-11	WD0626	6.00E-16
Candidatus *Neoehrlichia lotoris*	WP_084229825.1	8.00E-10	WD0626	5.00E-15

Table B-3. Primers used in this study.
All primers, names, and sequences are listed, along with the Y-chromosome FISH probe sequence that also has a 5' Cy5 tag.

Gene	Primer Name	Primer Sequence
WD0626 (*wmk*) native	WD0626_F	AATTGGCCTCTCTGCTAATGAGTG
	WD0626_R	CACGTCCTTGCTCATAGTTGCTTA
WD0626 (*wmk*) transgene	WD0626opt_F	TCCAGTGAGCTCCGAGAAGA
	WD0626opt_R	CCACGCGGGTAAACTTTGTC
WD0034 (control) native	WD0034_F	GGAAGAAACTTGCACACCACTTAC
	WD0034_R	TGCTCTCCGACCATCTGGATATTT
WD0034 (control) transgene	WD0034opt_F	TTAAGTACCCAGACGGACGC
	WD0034opt_R	TCCTTGTTGTCGGGATAGCG
WD0625 transgene	WD0625opt_F	CGCGAGATGGATGACCTGAA
	WD0625opt_R	CTCGCGCTCACTATGTCCAA
WD0508 transgene	WD0508opt_F	GACGTGCTGATCAAGAGCCT
	WD0508opt_R	TGCCCACTGTCTTCAGGATG
*Wolbachia gro*EL standard	groELstd_F	GGTGAGCAGTTGCAAGAAGC
	groELstd_R	AGATCTTCCATCTTGATTCC
*Wolbachia gro*EL	groEL_F	CTAAAGTGCTTAATGCTTCACCTTC
	groEL_R	CAACCTTTACTTCCTATTCTTG
WD0631 (*cif*A) native	WD0631_F	TGTGGTAGGGAAGGAAAGAGGAAA
	WD0631_R	ATTCCAAGGACCATCACCTACAGA
Rpl36 (*Drosophila*)	Rpl36_F	GTTTAATTCTCAAGTAACGTCATC
	Rpl36_R	TGTCCAACATCCTCACC
5'Cy5 FISH probe Y chromosome (Cheng et al 2016)	5'Cy5YChrome	AATACAATACAATACAATACAATACAATAC
Wolbachia 16S	Wolb_F	GAAGATAATGACGGTACTCAC
	Wolb_R3	GTCACTGATCCCACTTTAAATAAC
*w*Bif *wmk* homolog	wmk_wBif_F	AGGTTCGTGATACGGTGTGT
	wmk_wBif_R	ATCTGTGTACGCCCTCTTGC
*w*Bif *gro*EL	wBif_groEL_F	CGGGTTATAAGATTGCAGAAGGTG
	wBif_groEL_R	GAGATGCCACATCCAGCAATATTC
*w*Bif *cifA* homolog	wBif_CifA_F	GAGATGGCTTGTAGTTACTGTGTG
	wBif_CifA_R	GACCTTTCCTTCGAATGCCACC

Table B-4. Wmk protein homologs included in Fig. B-1b.
All homologs listed are those used for Fig. B-1b. Accession numbers for NCBI are included.

wmk homolog label	Strain	NCBI Accession	Reciprocal Best BLASTp to wMel	Reciprocal BLASTp E-value to Wmk
Wmk	*w*Mel	WP_010962718.1	WD0626	0
*w*Rec 0560	*w*Rec	WP_038198911.1	WD0626	0
*w*Inn (0626)	*w*Inn	MK873082	WD0626	0
*w*Bor (0626)	*w*Bor	MK873003	WD0626	9.E-163
*w*Mel WD0255	*w*Mel	WP_010962465.1	WD0255	2.E-153
*w*Mel WD0623	*w*Mel	WP_010962717.1	WD0623	2.E-157
*w*Inn (0623)	*w*Inn	MK873081	WD0255	4.E-111
*w*Bor (0623)	*w*Bor	MK873002	WD0255	2.E-155
*w*Mel WD0508	*w*Mel	WP_010962645.1	WD0508	5.E-108
*w*Mel WD0622	*w*Mel	WP_010962716.1	WD0622	3.E-76
*w*Inn (0622)	*w*Inn	MK873080	WD0622	6.E-50
*w*Bor (0622)	*w*Bor	MK873001	WD0622	3.E-73
*w*Bif (0626)	*w*Bif	MK873005	WD0626	6.E-26

Table B-5. *wmk* gene homologs included in Fig. B-1a.
All homologs listed are those that were included in the phylogeny from Fig. B-1a, or were excluded for indicated reasons. Accession numbers for NCBI are included.

wmk homolog label	Strain	Supergroup	NCBI Accession	Contig name	Genome/contig length	Homolog location	Locus Tag	Query coverage	E-value	% identity	
wmk	*w*Mel	A	AE017196.1		1267782	611371-612282	WD0626	100	0	100	
*w*Ha 02320	*w*Ha	A	CP003884.1		1295804	280755-281666	wHa_02320	100	0	99	
*w*Ri 005880	*w*Ri	A	CP001391.1		1445873	632500-633411	wRi_005880	100	0	99	
*w*Ri 010540	wRi	A	CP001391.1		1445873	1138959-1139870	wRi_010540	100	0	99	
*w*IncCu 02670	*w*Inc_Cu	A	CP011148.1		1267840	611424-612335	WG67_02670	100	0	99	
*w*Pip 0239	*w*PipPel	B	AM999887.1		1482455	247657-248570	WP0239	99	0	88	
*w*Au 0252	*w*Au	A	LK055284.1		1268461	250436-251356	WPWAU_0252	100	0	86	
*w*Au 0691	*w*Au	A	LK055284.1		1268461	666201-667055	WPWAU_0691	94	0	85	
*w*MelPop 03503	*w*MelPop	A	AQQE01000043.1	contig_00005_6	49467	26890-27801	WMELPOP_03503	100	0	100	
*w*Suzi 042	*w*Suzi	A	CAOU02000042.1	wsuzi2_contig042	1972	266-1177	N/A	100	0	99	
*w*Suzi 009	*w*Suzi	A	CAOU02000022.1	wsuzi2_contig009	57878	263-1174	N/A	100	0	99	
*w*Ana 0875	*w*Ana	A	AAGB01000100.1	gdan_354	2888	1216-2127	WwAna0875	100	0	99	
*w*Rec 0560	*w*Rec	A	JQAM01000018.1		11989	519-1430	wrec0560	100	0	99	
*w*Nfe 31	*w*Nfe	A	LYUY01000031.1	NODE_31	12399	703-1598	N/A	98	0	94	
*w*Npa 15a	*w*Npa	A	LYUX01000015.1	NODE_15	19352	7720-8615	N/A	98	0	94	
*w*Nfla 18a	*w*Nfla	A	LYUW01000018.1	NODE_18	19541	10921-11822	N/A	98	0	94	

*w*Nleu 18a	*w*Nleu	A	LYUV01000018.1	NODE_18	19173	7718-8613	N/A	98	0	94	
*w*Bol1b 0133	*w*Bol1-b	B	CAOH01000056.1	contig 01_7	10792	9781-10692	wBol1_0133	100	0	88	
*w*Npa 14	*w*Npa	A	LYUX01000014.1	NODE_14	19790	18117-19034	N/A	100	0	87	
*w*Nfe 23	*w*Nfe	A	LYUY01000023.1	NODE_23	14002	11199-12116	N/A	100	0	87	
*w*Nfla 78a	*w*Nfla	A	LYUW01000078.1	NODE_78	5269	876-1793	N/A	100	0	87	
*w*Nleu 8a	*w*Nleu	A	LYUV01000008.1	NODE_8	24767	18223-19140	N/A	100	0	87	
*w*Wil 404a	*w*Wil	A	AAQP01000017.1	TSC#14030-0811.24 1101007000404	9036	2974-3894	N/A	100	0	86	
*w*DacB 06900	*w*DacB	B	LSYY01000169.1	Contig_72	3902	1742-2647	TV41_06900	100	0	85	
*w*DacA 05595	*w*DacA	A	LSYX01000020.1	Contig_116	2049	23-908	TV42_05595	96	0	85	
*w*AlbB 49002	*w*AlbB	B	CAGB01000110.1	contig00334-1405	6182	1185-2093	WALBB_490002	98	0	85	
*w*CauB	*w*CauB	B	MK955149				N/A		0	88	
*w*Bor	*w*Bor	A	MK873001				N/A		0	85	
*w*Inn	*w*Inn	A	MK873080				N/A		0	85	
* Sequences identified during this study											
Not Included in Phylogeny											
***wmk* homolog label**	**Strain**	**Supergroup**	**NCBI Accession**	**Contig name**	**Genome/contig length**	**Homolog location**	**Locus Tag**	**Query coverage**	**E-value**	**% identity**	**Reason for exclusion**
*w*IncSM 02660	*w*Inc_SM	A	CP011149.1		1267664	611253-612163	WH35_02660	100	0	98	Disrupted ORF
*w*Sim 0298	*w*Sim	A	AAGC01000294.1	gdsi_178	1164	1-742	WwSim0298	81	0	100	Partial, end of contig
*w*Ana 0166	*w*Ana	A	AAGB01000245.1	gdan_78	1374	722-1374	WwAna0166	71	0	100	Partial, end of contig
*w*PipMol 01121	*w*PipMol	B	CTEH01000009.1	LargeContigsSCcontig000009	12073	3137-4050	WPM_01121	99	0	88	Flanked by gaps, problematic genome assembly
*w*PipJHB 1378	*w*PipJHB	B	ABZA01000018.1	contig_1290	1879	838-1751	C1A_1378	99	0	88	Problematic genome assembly
*w*Bol1b 0010	*w*Bol1-b	B	CAOH01000062.1	contig 18_3	4184	3181-4084	wBol1_0010	100	0	88	Disrupted ORF
*w*DacA 04625	*w*DacA	A	LSYX01000092.1	Contig_4	5785	1180-2095	TV42_04625	100	0	87	Disrupted ORF
*w*PipMol 01211	*w*PipMol	B	CTEH01000032.1	LargeContigsSCcontig000032	1541	578-1489	WPM_01211	100	0	86	Problematic genome assembly
*w*PipJHB 1294	*w*PipJHB	B	ABZA01000007.1	contig_1303	29919	25707-26618	C1A_1294	99	0	85	Problematic genome assembly

Table B-6. Amino acid similarity between Wmk and homologs in male-killing strains. Strains include all those that are currently sequenced.

Homolog	% aa similarity to *w*Mel Wmk
*w*Rec Wmk	99.70%
*w*Bol1b Wmk	85.90%
*w*Caub Wmk	83.30%
*w*Inn/*w*Bor Wmk	77.90%
*w*Bif Wmk	26.20%

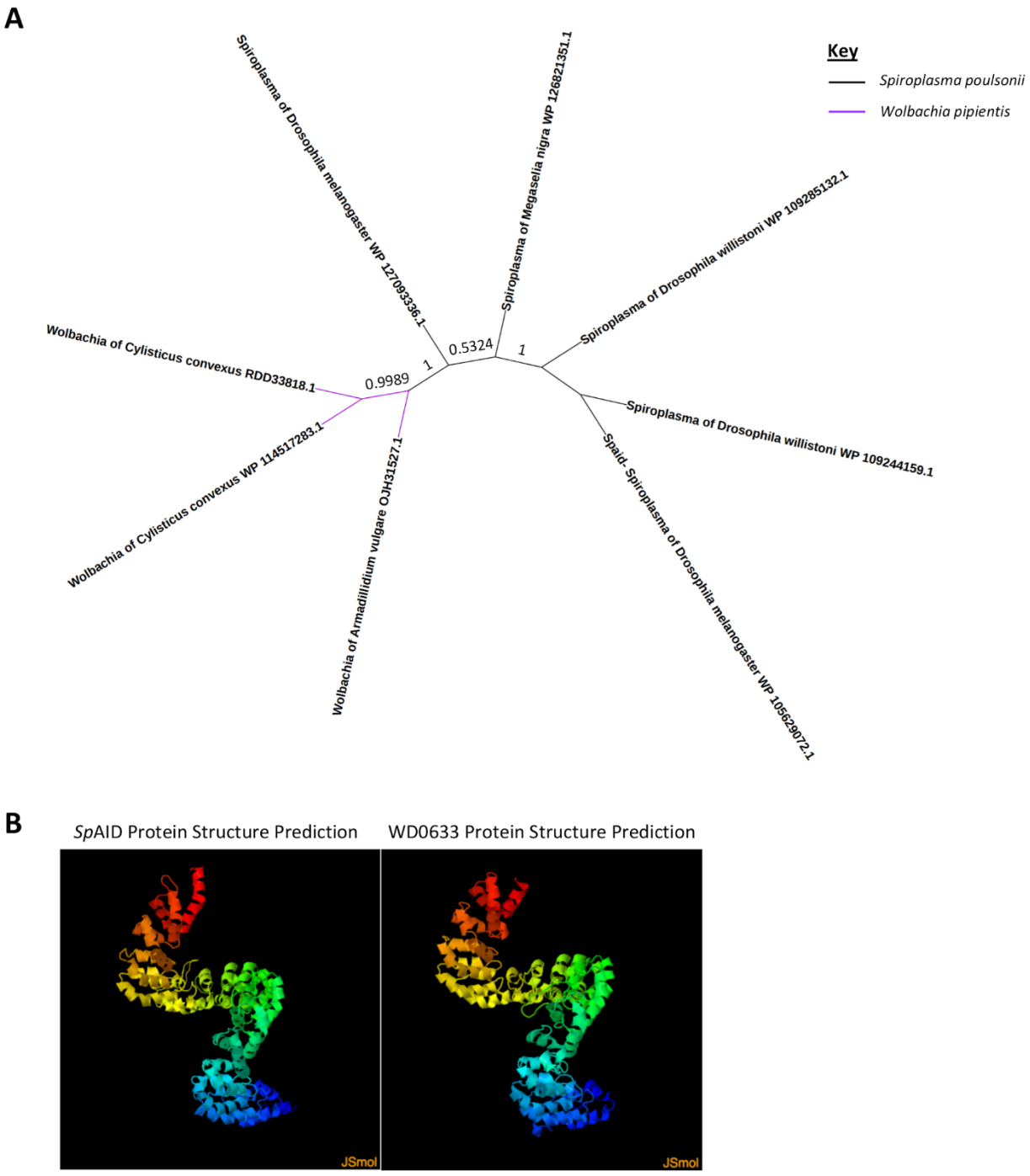


Figure C-1. *Sp*AID is similar in sequence and structure to *Wolbachia* proteins.
(a) A Bayesian phylogeny of the OTU domain of *Sp*AID and homologous sequences from NCBI based on a 54 aa-alignment using the cpRev model of evolution. Posterior probability values are shown on the branches. Purple branches correspond to *Wolbachia* sequences and green branches correspond to *Spiroplasma* branches. (b) Protein structure predictions from the Phyre2 web portal[466]. Images are of the most likely 3D structures of *Sp*AID and WD0633 determined by

Phyre2. *Sp*AID has 100% confidence, 54% coverage, and 9% sequence id to the top model based on the c4cj9A template (BurrH DNA-binding protein from *Burkholderia rhizoxinica* in apo form). WD0633 has 100% confidence, 58% coverage, and 10% id to the same model. Colors are in order of the rainbow from 5' to 3' end of the modeled portion of the genes.

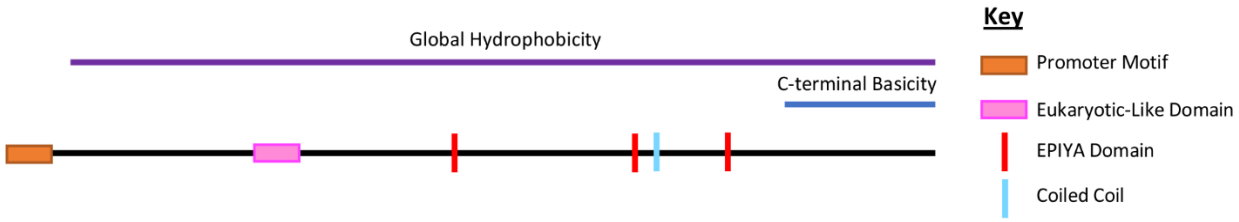


Figure C-2. Type IV secretion system (T4SS) motifs in WD0626.
Domains and motifs identified by the S4TE program are labeled[469]. C-terminal basicity refers to the entire C-terminal region, and Global Hydrophobicity refers to the entire protein. The black line indicates the protein sequence and other lines and boxes indicate T4SS effector-like features.

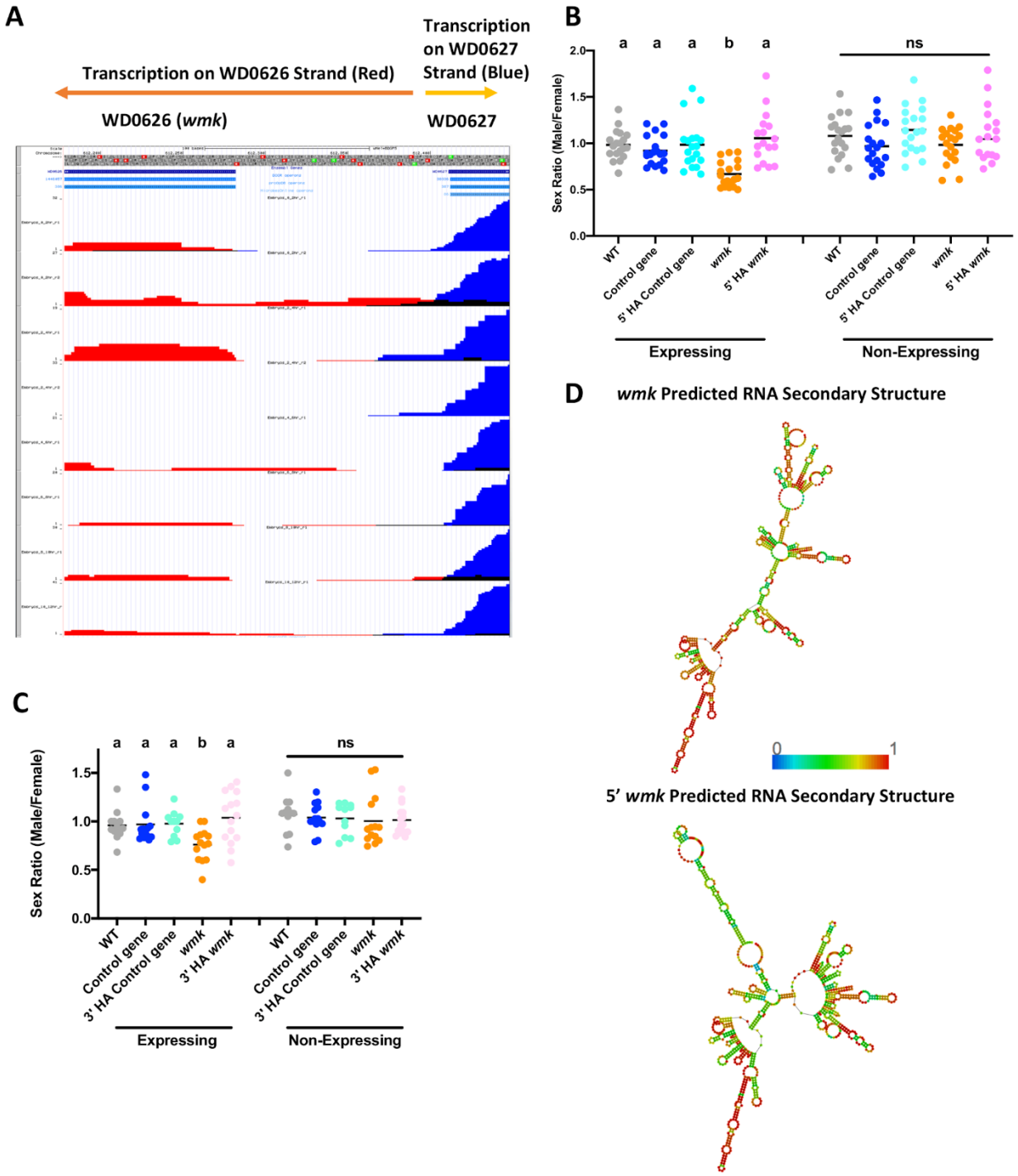


Figure C-3. Alternative transcripts of *wmk* may have different structures and functions. (a) Image of gene transcription of *w*Mel-infected *Drosophila melanogaster* embryos from the UCSC Genome Browser database[492]. Each row represents a sample of the indicated embryo age. Red indicates transcription on the same strand as *wmk*, and blue indicates transcription on the opposite strand with WD0627. Most samples have transcription upstream of the annotated *wmk* start codon. (b) Sex ratios of adult flies either expressing (*Act5c*-Gal4) or not expressing (CyO) the indicated genes, including the 5' HA tag. (c) Sex ratios of adult flies either expressing (*Act5c*-Gal4) or not expressing (CyO) the indicated genes, including the 3' HA tag. Each sample point

represents the adult offspring 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. (d) Predicted RNA secondary structures of *wmk* and *wmk* with the alternative start codon tested in Fig. 5-6c. The predicted structure was produced by the RNAfold web server[493], and represents the MFE secondary structure. The colors are indicative of base-pair probabilities with 0 probability represented in blue and high probability (1) represented in red.

Table C-1. List of *w*Mel genes scored by S4TE algorithm.

List of *w*Mel genes that scored above the threshold value in the S4TE algorithm for type IV secretion system effectors by locus tag and score. The entire *w*Mel genome was searched, and yielded 148 putative effectors out of 1195 total proteins and a threshold score of 72.

	Gene Locus Tag (WDXXXX)	Score
1	0633	246
2	0073	245
3	0514	245
4	0147	235
5	0286	203
6	0294	203
7	0385	200
8	0636	191
9	0346	173
10	0550	171
11	0438	169
12	0766	161
13	0285	160
14	1314	160
15	0466	156
16	0035	151
17	0291	151
18	0498	148
19	1228	144
20	0566	143
21	0365	138
22	0637	138
23	0024	137
24	0335	137
25	0513	137
26	0040	134

27	0221	134
28	0548	131
29	0596	131
30	0028	127
31	0464	127
32	0754	126
33	0224	124
34	0686	124
35	0696	124
36	0942	121
37	1237	121
38	0484	119
39	1144	119
40	0154	117
41	0317	117
42	0424	117
43	0069	115
44	0348	115
45	0131	114
46	0630	114
47	0835	114
48	1133	114
49	1277	114
50	0079	112
51	0247	112
52	0320	112
53	0928	112
54	1309	107
55	0632	105
56	0796	105
57	1199	105
58	0292	103
59	0231	102
60	0441	102
61	0462	102
62	0975	102
63	1245	102
64	0512	101
65	0549	101

66	1298	101
67	0706	99
68	0880	99
69	1212	99
70	0359	94
71	0465	94
72	0582	94
73	0609	94
74	0745	94
75	0839	94
76	0978	94
77	1082	94
78	1161	94
79	1173	94
80	0191	92
81	0371	92
82	0565	92
83	0776	92
84	1094	90
85	0445	89
86	0862	89
87	0060	87
88	0288	87
89	0382	87
90	0610	87
91	0634	87
92	0838	87
93	1174	87
94	0580	85
95	1160	85
96	0423	84
97	0485	84
98	0509	84
99	0631	84
100	0733	84
101	1039	84
102	1318	84
103	0026	83
104	0212	83

105	0213	82
106	0429	82
107	0576	82
108	0644	82
109	0649	82
110	0916	82
111	1171	82
112	1179	82
113	1202	82
114	1279	82
115	0198	81
116	0443	81
117	0539	80
118	1088	80
119	0170	79
120	0690	79
121	0798	79
122	0832	79
123	1223	79
124	0906	78
125	0353	77
126	0783	77
127	0797	75
128	0824	75
129	1041	75
130	1213	75
131	0223	74
132	1051	74
133	1111	74
134	0557	73
135	0034	72
136	0251	72
137	0338	72
138	0389	72
139	0413	72
140	0594	72
141	0602	72
142	0763	72
143	0764	72

144	0830	72
145	0892	72
146	0897	72
147	0981	72
148	1014	72

Table C-2. List of alternative start codons to *wmk* and homologs.
List of alternative start codons in the same frame as the annotated codon in *w*Mel and male-killer genomes of *Drosophila* hosts. A star symbol (*) indicates the primary start codon annotated by Geneious. Total length includes the length of the stop codon.

Genome	Codon	Additional Length (Amino Acids)	Total length (Amino Acids)
*w*Mel	ATG*	0	304
*w*Mel	ATA	2	306
*w*Mel	ATA	3	307
*w*Mel	TTG	9	313
*w*Mel	ATC	16	320
*w*Mel	ATA	20	324
*w*Rec	ATG*	0	304
*w*Rec	ATA	2	306
*w*Rec	ATA	3	307
*w*Rec	TTG	9	313
*w*Rec	ATC	16	320
*w*Rec	ATA	20	324
*w*Ri	ATT*	0	304
*w*Ri	ATG	6	310
*w*Ri	ATT	34	338
*w*Ri	ATG	35	339
*w*Ha	ATG*	0	304
*w*Ha	ATA	2	306
*w*Ha	ATA	3	307

*w*Ha	TTG	9	313
*w*Ha	ATC	16	320
*w*Ha	ATA	20	324
*w*Bif	ATG*	0	257
*w*Bif	ATA	2	259
*w*Inn	ATA*	0	313
*w*Inn	GTG	14	327
*w*Bor	GTG*	0	313
*w*Bol1-b	ATG*	0	304
*w*Bol1-b	ATA	2	306
*w*Bol1-b	ATA	3	307
*w*Bol1-b	TTG	9	313
*w*Bol1-b	ATA	17	321
*w*Au	ATG*	0	307
*w*Au	ATA	2	309
*w*Au	ATA	3	310
*w*Au	ATA	8	315
*w*Au	TTG	9	316
*w*Au	ATT	21	328

Appendix D. Chapter VI Supplementary Information

Table D-1. 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	GGA TTA ACG CGG TCT AAG CAT GTG TAA CTG
Msl-2_R	GTA TGC CGT CTG GGC CAT GAT G

Appendix E. Microbial misandry: Discovery of a *Spiroplasma* male-killing toxin

Contributing Authors: Jessamyn I. Perlmutter and Seth R. Bordenstein

Summary

Bacteria-induced male killing evolved to enhance maternal transmission of the bacteria. Despite significance to arthropod evolution and vector control potential, the genetic basis behind this selfish microbial trait has remained mysterious. In recent work, Harumoto & Lemaitre (2018)[248] describe a *Spiroplasma poulsonii* male-killing toxin sought for over half a century.

Main Text

Bacterial endosymbionts of arthropods represent some of the most widespread microbes in the animal kingdom. These intracellular bacteria primarily inhabit the gonads of their arthropod hosts and are cytoplasmically inherited from mother to egg via the ovarian stem cells, similar to mitochondria. As these bacteria depend on their host for viability and transmission, they employ various strategies to ensure their transovarial passage to the next generation. In some cases, the bacteria facilitate their spread by conferring fitness benefits to the host. For example, they may increase the fecundity of females, which in turn boosts the frequency of bacteria in the next generation. In other cases, these microbes cunningly manipulate host reproduction, facilitating their own spread in what is termed "reproductive parasitism". These manipulations span five known phenotypes: (i) Cytoplasmic Incompatibility (CI), or death of offspring with uninfected mothers and infected fathers, (ii) Parthenogenesis, or female cloning of the infected matriline, (iii) Feminization, or development of genetic males into morphological and reproductive females, (iv) Meiotic drive, or exclusion of the maternally-inherited sex chromosome required to form males, and (v) Male Killing, or selective death of infected, male offspring. All phenotypes except CI distort the host sex ratio in favor of infected females, the transmitting sex of the bacteria, because males are an evolutionary dead end for the bacteria. Reproductive parasitism is so successful that it has arisen in several bacterial lineages, including *Spiroplasma, Wolbachia, Rickettsia, Cardinium,* and *Arsenophonus,* in approximately half of all arthropod species worldwide.

Some of the earliest studies that focused on male killing described a heritable "sex-ratio" factor [494] that was later named *Spiroplasma poulsonii*. Ensuing research on *Spiroplasma* and other

male killers has given a greater understanding of the evolution of male killing. At first glance, killing males seems a counterintuitive survival strategy, but the key concept is that what benefits the infected females also benefits the bacteria. Male killing halves sibling competition in cases of limited resources, whereby infected sisters no longer compete with their infected brothers for those resources. In some cases, infected sisters cannibalize their dead brothers, giving them extra nutrients. Of course, high levels of male killing must be balanced by an influx of males from nearby populations or incomplete bacterial transmission, otherwise the sexually-reproducing population will go extinct. Male killing can also impose strong selection on host genetic variation to counter the sex-specific lethality by altering mating strategies, clutch sizes, and bacterial transmission. Indeed, host resistance to male killing is reported [329]. Population modelling further shows that male killing may be an effective method of reducing population sizes of arthropod pests or vectors of disease [495].

Despite longstanding interest, the microbial genetic basis of male killing has remained a mystery. Progress has been hindered by the genetic intractability and difficulty in culturing the primarily intracellular bacteria. Researchers have recently turned to comparative genomics and heterologous gene expression techniques, such as the bipartite Gal4-UAS system in *Drosophila*, to study the genetics of reproductive parasitism [158,429,496]. The Gal4-UAS system enables spatiotemporal regulation of transgene expression and has proven particularly useful for expression of bacterial or phage genes in *Drosophila*. This system provides inroads to assaying putative reproductive parasitism candidate genes, despite bacterial impediments. For male killing, work until now has focused on correlated host defects. Previously, Harumoto et al. discovered that *Spiroplasma* male killing induces host apoptosis and male-specific DNA damage via the male-specific lethal (MSL) complex that mediates dosage compensation in males [360].

In a major advance for the field, Harumoto and Lemaitre have now shed light on the microbial genetic basis of *Spiroplasma*-induced male killing [497]. Serendipitously, a strain of *Spiroplasma* that exhibited incomplete male killing in *Drosophila melanogaster* arose in the lab. They took advantage of this phenotypic variation by sequencing the genome and comparing it to the genome of a complete male-killing strain. A candidate gene, Spaid (*S. poulsonii* androicidin, after the fabled *Spiroplasma* male-killing toxin named long ago), encodes a 1065 amino acid protein with ankyrin repeats and an ovarian tumor (OTU) deubiquitinase domain. In the strain with incomplete male killing, Spaid has an 828-base pair deletion. This truncated Spaid protein lacks

the C-terminal hydrophobic region with no predicted domains, suggesting uncharacterized regions of the gene contribute to male killing. As for the other domains in Spaid, ankyrins have been hypothesized to be involved in reproductive parasitism phenotypes. Ankyrin domains mediate protein interactions, are common in eukaryotes, and are unusually enriched in obligate intracellular bacteria, which suggests a potential role in interactions with host cells [496,498]. The other Spaid domain, an OTU deubiquitinase, is also present in one of the CI-causing genes, suggesting it may play a role in more than one case of reproductive parasitism [429]. Notably, a *Wolbachia* gene with an ankyrin repeat and OTU deubiquitinase domain, WD0633, previously showed no phenotype upon transgenic expression in *D. melanogaster* [496]. Coupled with the absence of Spaid homologs in other male-killing bacteria, this suggests that male killing may have independent origins in several lineages, as suggested previously [499]. Notably, Spaid is found on a plasmid in *Spiroplasma*. This is similar to *cytoplasmic incompatibility factors A* and *B* (*cifA* and *cifB*), which have previously been defined as genes causing *Wolbachia*-induced CI and are found in prophage WO in the bacterium [158]. Thus, mobile elements may play a significant role in the origins and spread of selfish endosymbiont traits.

To functionally evaluate Spaid, Harumoto and Lemaitre transgenically expressed the gene in uninfected *D. melanogaster* via the Gal4-UAS expression system. The basic idea is that if transgenic expression recapitulates *Spiroplasma*-based male killing, it is likely that the gene contributes to the male-killing phenotype. Indeed, expression of Spaid selectively kills larval male offspring, but leaves females unharmed. Larval death is later in development than wild type *Spiroplasma*-induced male killing that occurs during embryogenesis, which suggests either other genes or expression patterns may be required to induce the full phenotype in the normal embryonic stage. However, the authors show transgenic expression recapitulates the cytological defects of the phenotype. Apoptosis is significantly higher in males and is induced in females that artificially express the MSL complex. Spaid, therefore, appears to depend on the MSL complex to kill its host. DNA damage is increased in males, and the localization of the damage overlaps with the MSL complex. Spaid was also found significantly more often in the nucleus at the same location as the MSL complex.

To investigate the essential regions necessary for male killing, transgenic expression of Spaid with a deletion of either the ankyrin or OTU domain was performed. Expression of Spaid without the ankyrin domain does not kill males, and it localizes in the nucleus. Expression of Spaid

without the OTU domain did kill some males, but in later pupal development, and was not found as concentrated in the nucleus. However, when Spaid with the OTU domain deletion was transgenically expressed with a weaker Gal4 driver that produces less transcript, the male killing phenotype was lost. Taken together, the authors suggest a model of *Spiroplasma*-based male killing whereby the ankyrin domain promotes interaction with the MSL complex to mediate chromatin modifications and ensuing male death, while the OTU domain contributes to nuclear localization.

Following the identification of the *Spiroplasma* MSRO (*melanogaster* sex ratio organism) male-killing gene, several fundamental questions remain. First, what host factor(s) interact with Spaid, and what molecular events occur that cause male-specific death? This question is critical to understand the specific effects on hosts and the potential host range, likelihood of host resistance, and other host factors that may impact vector control efforts utilizing male killing. Second, what is the breadth of this finding? *Spiroplasma* kills several *Drosophila* hosts, but others are affected as well including butterflies and ladybugs. As *Drosophila* has a unique dosage compensation mechanism, it will be interesting to determine if these findings apply in all hosts, or if there are differences in mechanism. It is possible that various *Spiroplasma* strains independently developed several unique male-killing toxins or that Spaid has diverse homologs that have adapted to kill males of distantly-related hosts. It will be important to compare and test putative Spaid homologs in other male killing strains, which will require sequencing these genomes. Finally, how does Spaid compare to other microbial male-killing genes? Independent evolution of male killing in multiple bacterial lineages appears likely as there are no obvious Spaid homologs in sequenced strains of other male killers. It will be exciting to compare and contrast the male killing genes and mechanisms evolved by various microbes and to assess their breadth, diversity, and potential applications to vector control.

Appendix F. List of Publications

LePage, DP*, Metcalf, JA*, Bordenstein, SR, On, J, **Perlmutter, JI,** Shropshire, JD, Layton, EM, Funkhouser-Jones, LJ, Beckmann, JF, & Bordenstein, SR. (2017). Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. *Nature, 543* (7644), 243. *Co-first authors

Perlmutter, JI & Bordenstein, SR. (2018). Microbial misandry: Discovery of a *Spiroplasma* male-killing toxin. (Preview). *Cell Host & Microbe*, *23* (6), 689-690.

Perlmutter, JI, Bordenstein, SR, Unckless, RL, LePage, DP, Metcalf, JA, Hill, T, Martinez, J, Jiggins, FM, & Bordenstein, SR. (2019). The phage gene *wmk* is a candidate for male killing by a bacterial endosymbiont. *PLoS Pathogens, 15* (9), e1007936.

Perlmutter, JI & Bordenstein, SR. (2020). Microorganisms in the reproductive tissues of arthropods. (Review). *Nature Reviews Microbiology*, 1-15.

Perlmutter, JI, Meyers, JE, & Bordenstein SR. (2020). Transgenic testing does not support a role for additional candidate genes in *Wolbachia* male killing or cytoplasmic incompatibility. *mSystems, 5* (1).

Perlmutter, JI, Meyers, JE, & Bordenstein SR. Assessment of transgenic homologs of a male-killing gene candidate. *In prep (Chapter 6).*

Perlmutter, JI, Meyers, JE, & Bordenstein, SR. Microbe-mediated male killing. (Review). *In prep (Chapter 2).*