EVOLUTIONARY AND FUNCTIONAL STUDIES OF *WOLBACHIA PIPIENTIS* AND ITS PHAGE

By

Jason A. Metcalf

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

December, 2014

Nashville, Tennessee

Approved:

Professor Seth R. Bordenstein

Professor Patrick Abbot

Professor Julián F. Hillyer

Professor Christopher R. Aiken

Professor Maureen A. Gannon

DEDICATION

To my wife and partner in all things, Elise Pfaltzgraff, and to my parents, for their love and support in my every endeavor.

ACKNOWLEDGEMENTS

I am incredibly grateful to the many mentors and friends who have supported me through the years. I would not be the person I am today without them. Thank you to my parents, Jeffrey and Carrie Metcalf, who stimulated my interest in science from an early age and have been there for me ever since, and to my high school biology teacher, Alta Krajec, who solidified that interest into a path in higher education. I am extremely appreciative of my undergraduate mentor, Dr. Katherine Ponder, who took me in when I knew nothing, and developed my scientific skills over the course of five years at Washington University in St. Louis. I would also like to thank my thesis committee, both past and present, Dr. Patrick Abbot, Dr. Julián Hillyer, Dr. Christopher Aiken, Dr. Maureen Gannon, Dr. Laura Lee, and Dr. Joshua Gamse. Your insight and advice has been invaluable. I thank my thesis mentor, Dr. Seth Bordenstein, for guiding me through this monumental endeavor. Thank you for being there with words of advice and encouragement when things were not going well, for celebrating with me when they were, for listening to and even fostering my crazy ideas, and for inserting a voice of reason when I needed it.

I have been fortunate to have a multitude of fantastic collaborators and colleagues who put significant effort into the work presented here. Specifically, I would like to thank Lisa Funkhouser-Jones, Dr. Kristen Brileya, Dr. Anna-Louise Reysenbach, Minhee Jo, Sarah Bordenstein, Dr. John Jaenike, Daniel LePage, William Martin, and Dr. Nicholas Reiter, all of who have contributed as coauthors to the research herein. I am also very

appreciative to the many people who have provided technical support and advice for my experiments, including Dr. Antonis Rokas, Dr. Jennifer Wisecaver, Dr. Kristin Jernigan, Rini Pauly, Dr. Bethany Kent, Dr. Robert Brucker, Joey Simmons, Teddy van Opstal, Andrew Brooks, Ashley Saulsberry, Christine Sislak, Annie Lindgren, Alex Hirschi, Dr. Julie Dunning Hotopp, and Dr. Barton Slatko. Many thanks as well to the Vanderbilt Antibody and Protein Core, the Vanderbilt Imaging Core, and Dr. James Patton for access to equipment and reagents.

I must also express my thanks to the numerous friends and colleagues who never had a hand in the science, but were always willing to hear about it and support me when the going got tough. Finally and most importantly, I thank my wife, Elise Pfaltzgraff, without whose love and support I would be lost. You are my greatest discovery, the love of my life, and my best friend. I can't wait to see where life takes us from here.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF TABLES viii

LIST OF FIGURES ix

LIST OF ABBREVIATIONS xi

Chapter

I. INTRODUCTION 1

Introduction to *Wolbachia pipientis* 1
Phage WO 2
Wolbachia and horizontal gene transfer 3
Wolbachia in *Drosophila recens* 4
Mediators of cytoplasmic incompatibility 6
Wolbachia evolution during host switching 6
Conclusions and future directions 7

II. THE COMPLEXITY OF VIRUS SYSTEMS: THE CASE OF ENDOSYMBIONTS. 8

Abstract 8
Introduction 8
Prevalence of phages in endosymbionts 10
Evolution of WO 11
Involvement of WO in reproductive parasitism 16
Applications of WO 19
Conclusions 20

III. PARALLEL HORIZONTAL TRANSFER AND FUNCTION OF AN ANTIBACTERIAL GENE ACROSS THE TREE OF LIFE 21

Abstract 21
Introduction 22
Results and Discussion 24
A Bacterial GH25 Muramidase Is Present in All Domains of Life 24
Non-bacterial GH25 Muramidases Arose From HGT 32
A. boonei GH25 Muramidase Is Antibacterial 39
Conclusions 48

Materials and methods 51
PCR and sequencing *51*
Lysozyme cloning and purification *53*
Antibacterial assays *55*
A. boonei cultures *55*

IV. RECENT GENOME REDUCTION OF *WOLBACHIA* IN *DROSOPHILA RECENS* TARGETS PHAGE WO AND NARROWS CANDIDATES FOR REPRODUCTIVE PARASITISM 57

Abstract 57
Introduction 58
Materials & Methods 60
Wolbachia strain relatedness *61*
Genome sequencing and assembly *61*
Annotation and comparative genomics *62*
Results 63
Taxonomy of wRec *63*
Genome features of wRec with targeted reduction of prophage WO *64*
Prophage WO relics in the genome *71*
Discussion 73
Divergence and genome reduction in wRec *73*
The phage WO hypothesis to explain reproductive parasitism *76*
WO host adsorption genes *78*
Future Studies *79*

V. *WOLBACHIA*-INDUCED CYTOPLASMIC INCOMPATIBILITY MAY BE CAUSED BY PHAGE-ENCODED VIRULENCE FACTORS 80

Abstract 80
Introduction 81
Criterion #1: Presence/absence in CI strains *82*
Criterion #2: Expression in gonads *83*
Criterion #3: Secretion *83*
Criterion #4: Previous candidate testing *84*
Finalizing candidates: function, toxin-antitoxin systems, and bidirectional incompatibility *84*
Materials and Methods 86
Candidate selection *86*
Candidate gene expression *87*
Transgenic Drosophila *88*
CI and rescue crosses *89*
Results and Discussion 90
Selection of candidate genes for further evaluation *90*
Bidirectional compatibility *93*
Candidate gene expression patterns *95*

Functional tests of CI *97*
Conclusions *101*

VI. MICROEVOLUTION AND INFECTION DYNAMICS OF *WOLBACHIA PIPIENTIS* AFTER HOST SWITCHING IN CELL CULTURE 103

Abstract 103
Introduction 104
Materials & Methods 106
Fly and cell lines *106*
Whole genome sequencing *106*
wMel culture experiments *107*
Results 109
Minimal genomic changes following introduction to cell culture *109*
wMel-RML12 mutation is in a conserved region of RNase P protein *110*
Infection dynamics during wMel host switching *112*
Discussion 114
Minimal genetic change associated with host switching *114*
Host-dependence of cell culture infection progression *116*

VII. CONCLUSIONS AND FUTURE DIRECTIONS. 117

Phage WO mechanisms of lysis 117
WO lysis candidates *118*
WO lysozyme is likely not solely responsible for Wolbachia lysis *120*
Proposed mechanism of WO lysis *122*
Antibacterial activity of WO lysozyme homologs 123
Mechanism of cytoplasmic incompatibility 125
Wolbachia adaptation to new hosts 127
Conclusions 128

REFERENCES 129

APPENDIX A. *W*MEL GENES MATCHING CI CANDIDATE CRITERIA 149

LIST OF TABLES

Table III-1. Field samples tested for presence of lysozyme gene. 27

Table III-2. Primers used in this study. 28

Table III-3. Bacterial strains used in antibacterial assays. 44

Table IV-1. *w*Rec sequencing and genome statistics. 66

Table IV-2. *w*Mel genes with less than 98% nucleotide identity to their orthologs in *w*Rec. 67

Table IV-3. *w*Mel genes with no homologs in *w*Rec. 69

Table V-1. Primers used in RT-qPCR. 88

Table V-2. *w*Mel genes matching two or more of CI candidate criteria #1-3 but not criterion #4. 90

Table VI-1. Primers used in this study. 108

Table VI-2. Genetic variation before and after *w*Mel host-switching. 111

Table VII-1. Proposed CI experiment crosses. 127

LIST OF FIGURES

Figure I-1. Multiple levels of symbiosis involving *Wolbachia*. 3

Figure I-2. Two major *Wolbachia*-induced reproductive manipulations........................... 5

Figure II-1. Effects of microbial ecology on exposure to phage gene pools. 9

Figure II-2. WO particle and genome structure. 12

Figure II-3. Evolution of bacteriophages in endosymbionts and free-living bacteria. 14

Figure II-4. Examples of gene flow between WO, *Wolbachia*, and insects. 15

Figure III-1. Presence of HGT lysozyme genes in field samples. 26

Figure III-2. PCR amplifications testing genomic integration with primers within and outside of lysozyme genes. 29

Figure III-3. Architecture of HGT candidates and surrounding genes. 30

Figure III-4. Protein phylogeny of neighboring genes to transferred lysozymes. 31

Figure III-5. Phylogeny of GH25 muramidase. 34

Figure III-6. Protein phylogeny of *A. oryzae* GH25 muramidase and relatives. 36

Figure III-7. DNA phylogeny of *A. oryzae* GH25 muramidase and relatives. 37

Figure III-8. Comparison of GH25 muramidase and rRNA divergence. 39

Figure III-9. Conservation of *A. boonei* GH25 muramidase domain. 40

Figure III-10. Lysozyme purifications. 41

Figure III-11. Antibacterial action of *A. boonei* GH25 muramidase domain against Firmicutes. 42

Figure III-12. Antibacterial test of *A. boonei* GH25 muramidase on additional bacteria. 43

Figure III-13. *E. coli* death following full-length *A. boonei* lysozyme expression. 45

Figure III-14. Lysozyme expression and relative fitness during *A. boonei* and *M. lauensis* coculture. 47

Figure IV-1. WO phage is present in all sequenced supergroup (A) and (B) *Wolbachia* strains. 64

Figure IV-2. *w*Rec genome comparison to *w*Mel. 70

Figure IV-3. Within-scaffold *w*Rec synteny compared to *w*Mel. 71

Figure IV-4. Number of phage genes in *w*Rec and its relatives. 73

Figure V-1. Identity matrix of WD0626 homologs in wMel and its relatives. 94

Figure V-2. Identity matrix for WD0631 homologs in relatives of wMel. 94

Figure V-3. WD0631 homologs in wPip Pel (top) and wPip Mol (bottom). 95

Figure V-4. Expression of CI candidate effectors in reproductive tissues. 96

Figure V-5. Hatch rate assays for *D. melanogaster* expressing either WD0625 or WD0626. 98

Figure V-6. Hatch rate assays for *D. melanogaster* expressing the indicated transgenes. 100

Figure VI-1. Conservation of WD_0200, a putative RNase P protein component, between sequenced strains of *Wolbachia*. 112

Figure VI-2. *w*Mel infection dynamics after switching or maintaining eukaryotic hosts. 114

Figure VII-1. Testing WORiA lysozyme activity in vitro. 121

Figure VII-2. Testing WORiA lysozyme effect on *Wolbachia* titer *in vivo*. 122

Figure VII-3. Proposed mechanism of *Wolbachia* lysis by WO. 123

Figure VII-4. Proposed *A. boonei* lysozyme constructs. 124

Figure VII-5. Insolubility of 6x-His tagged horizontally transferred lysozymes. 125

Figure VII-6. Solubility of recombinant wild type *w*Mel WD0200. 128

LIST OF ABBREVIATIONS

aa	Amino acid
ANOVA	Analysis of variance
ATTC	American type culture collection
bp	Base pairs
BLAST	Basic logic alignment search tool
BSA	Bovine serum albumin
C	Centigrade
CEWL	Chicken egg white lysozyme
CFP	Cyan fluorescent protein
CI	Cytoplasmic incompatibility
CloVR	Cloud virtual resource
DIAG	Data intensive academic grid
DIP	Deletion/insertion polymorphism
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
DUF	Domain of unknown function
GH2	Glycosyl hydrolase 2
GH25	Glycosyl hydrolase 25
GST	Glutathione S-transferase
HGT	Horizontal gene transfer
K_a	Number of nonsynonymous substitutions per nonsynonymous site
kb	Kilobase
K_s	Number of synonymous substitutions per synonymous site
MBP	Mannose-binding protein
mg	Milligram
μg	Microgram
mL	Milliliter
μL	Microliter
MLST	Multilocus sequence typing
MS	Mass spectrometry
OTU	Ovarian tumor
PAZ	Piwi / Argonaut / Zwille
PCR	Polymerase chain reaction
PLA2	Phospholipase A2
pTIGER	Plasmid for targeted integration, germline expression, UAS regulated
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RT-qPCR	Reverse transcriptase quantitative polymerase chain reaction
S4TE	Searching algorithm for type IV effector proteins
SDS	Sodium dodecyl sulfate
SEM	Standard error of the mean

SH3	Src-homology domain 3
SH-test	Shimodaira-Hasegawa test
siRNA	Small interfering RNA
SNP	Single nucleotide polymorphism
sp	Spanin
SUMO	Small ubiquitin-related modifier
T4SS	Type 4 secretion system
tRNA	Transfer ribonucleic acid
UAS	Upstream activation sequence

CHAPTER I. INTRODUCTION

Introduction to *Wolbachia pipientis*

Perhaps no other infection on the planet is more pervasive than that of *Wolbachia pipientis*, an obligate intracellular α-proteobacterium infecting nearly half of all arthropods (Zug and Hammerstein, 2012) and filarial nematodes (Ferri et al., 2011). *Wolbachia* was first detected in the mosquito, *Culex pipiens*, by Hertig and Wolbach in the 1920's (Hertig and Wolbach, 1924), and although initially regarded as a microbiological oddity, the basic and applied importance of *Wolbachia* has expanded dramatically in recent years after it was discovered that *Wolbachia* is involved in complex manipulations of its hosts (Yen and Barr, 1971) with practical applications to pest and disease vector control. Investigations into basic *Wolbachia* biology have revealed much about the evolution of intracellular bacteria (Moran et al., 2008) and their associated mobile genetic elements (Newton and Bordenstein, 2011, Kent et al., 2011a), as well as mechanisms of vertical transmission of endosymbionts (Toomey et al., 2013, Fast et al., 2011), modes of bacterial manipulation of their hosts (Werren et al., 2008), and methods of countering pathogens of major human health concern (Walker et al., 2011, Moreira et al., 2009). Despite these advances, a number of important biological questions remain regarding *Wolbachia* evolution and host interactions. In this thesis I describe progress made towards answering a few of those questions as described below.

Phage WO

Wolbachia is a central player in a multi-level symbiosis ranging from animals to insects to viruses (Figure I-1). In arthropods, *Wolbachia* is a reproductive parasite, manipulating the procreation of its hosts to favor the success of infected females who will pass down the bacteria to her offspring (Werren et al., 2008). *Wolbachia*'s insect host may also be parasitic, as is the case in *Nasonia* parasitoid wasps (Bordenstein et al., 2001), which lay their eggs in the pupae of blow flies that may in turn add another level of parasitism by populating the wounds of higher animals (Francesconi and Lupi, 2012). This hyperparasitism extends down the taxonomic ladder as well, as arthropod *Wolbachia* strains are themselves infected by a temperate bacteriophage named WO (Kent and Bordenstein, 2010). Once thought to be a phenomenon of free-living bacteria exclusively, we now know that intracellular bacteria such as *Wolbachia* may be extensively infected with bacteriophages (Bordenstein and Reznikoff, 2005). To date, all sequenced *Wolbachia* strains infecting arthropods have contained prophage sequences of WO, ranging from relatively short regions consisting of a few genes to up to five phage haplotypes encompassing 21% of the *Wolbachia* genome and including one or more prophages capable of producing active virions that lyse the *Wolbachia* cell (Kent et al., 2011a). Given the prevalence of WO in *Wolbachia* and the commonality of prophage-encoded virulence factors (Boyd, 2012), it is tempting to speculate that WO may somehow be involved in *Wolbachia*'s manipulation of its hosts (Saridaki et al., 2011, Sanogo and Dobson, 2006), an idea that will be explored further in Chapters IV and V. The recent literature regarding phage WO is reviewed in Chapter II, with a focus on

endosymbiont phage prevalence, WO evolution, and the tripartite interactions between WO, *Wolbachia*, and insects.

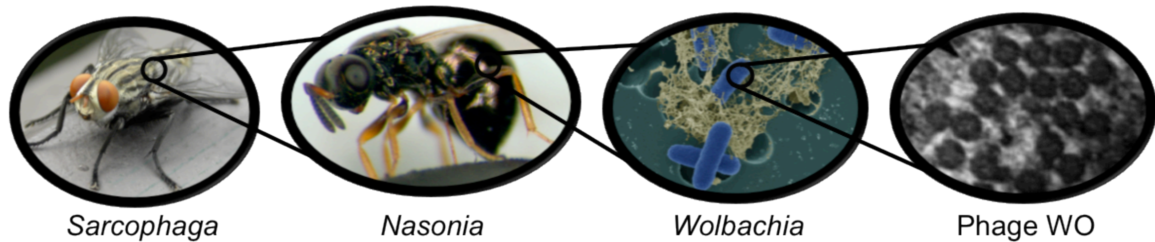


Figure I-1. Multiple levels of symbiosis involving *Wolbachia*. Blow flies of the genus *Sarcophaga* are parasitized by *Nasonia* wasps, which in turn are infected by *Wolbachia*. *Wolbachia* is itself infected by a temperate bacteriophage named WO. Image credits: Wikimedia commons, Sarah Bordenstein, and Robert Brucker.

Wolbachia and horizontal gene transfer

Horizontal gene transfer (HGT), the movement of genetic information between two unrelated species rather than by vertical descent, has long been recognized as a major player in prokaryote evolution (Popa and Dagan, 2011). More recently, it has become apparent that HGT between different domains of life may also be an important factor in the evolution of multicellular organisms (Dunning Hotopp, 2011, Keeling and Palmer, 2008). *Wolbachia* has been responsible for some of the most abundant interdomain HGT transfer events recognized (Dunning Hotopp et al., 2007), with anywhere from a few genes up to nearly the entire *Wolbachia* genome being transferred to its invertebrate host. The functional consequences of these transfers remain somewhat unclear (Dunning Hotopp, 2011), but a few selective advantages have been speculated from transferred gene function and expression patterns, such as *de novo* nucleotide synthesis (McNulty et

al., 2012) and regulation of nutritional symbionts (Nikoh et al., 2010). Interestingly, it appears that *Wolbachia* may be just a small part of an elaborate series of horizontal transfer events across the tree of life involving a phage WO gene with potential antibacterial function, a lysozyme in the glycosyl hydrolase (GH) 25 muramidase family. This gene has been transferred from diverse bacteria to ecologically associated plant fungus, insect, and archaea species. The evolution and function of this transferred gene, with a focus on the gene in the archaeon, will be discussed in Chapter III.

Wolbachia* in *Drosophila recens

Wolbachia is capable of several complex host manipulations that favor its spread in a population (Werren et al., 2008). One of those manipulations is male killing, wherein *Wolbachia* causes the majority of males in a brood to die, leaving a greater availability of resources to the females that will pass on their infection to their offspring (Hurst and Jiggins, 2000). A second reproductive manipulation is called cytoplasmic incompatibility (CI), in which crosses between *Wolbachia*-infected males and uninfected females are mostly inviable (Engelstadter and Telschow, 2009). Infected females however, have a selective advantage because they can rescue this inviability when mating with infected males, as well as successfully mate with uninfected males, passing on *Wolbachia* in both cases (Figure I-2). These manipulations and others have contributed greatly to *Wolbachia*'s evolutionary success across diverse arthropods, but the mechanisms behind these phenomena are poorly understood (Serbus et al., 2008). Interestingly, some strains of *Wolbachia*, such as the *w*Rec strain naturally infecting *Drosophila recens*, can cause

either male killing or CI, depending on the host in which it resides (Jaenike, 2007). *w*Rec is also of interest because it is one of the very few arthropod *Wolbachia* strains that appeared to lack phage WO by single gene PCR screen (Bordenstein and Wernegreen, 2004), making it an important test case for the hypothesis that WO may be involved in reproductive manipulations. Thus, to enable genomic, transcriptomic, and proteomic approaches that could elucidate mechanisms of multipotent reproductive manipulations and to investigate the phage WO anomaly, we sequenced the genome of *w*Rec and compared it to related *Wolbachia* strains. The outcome of this sequencing and analysis is reported in Chapter IV.

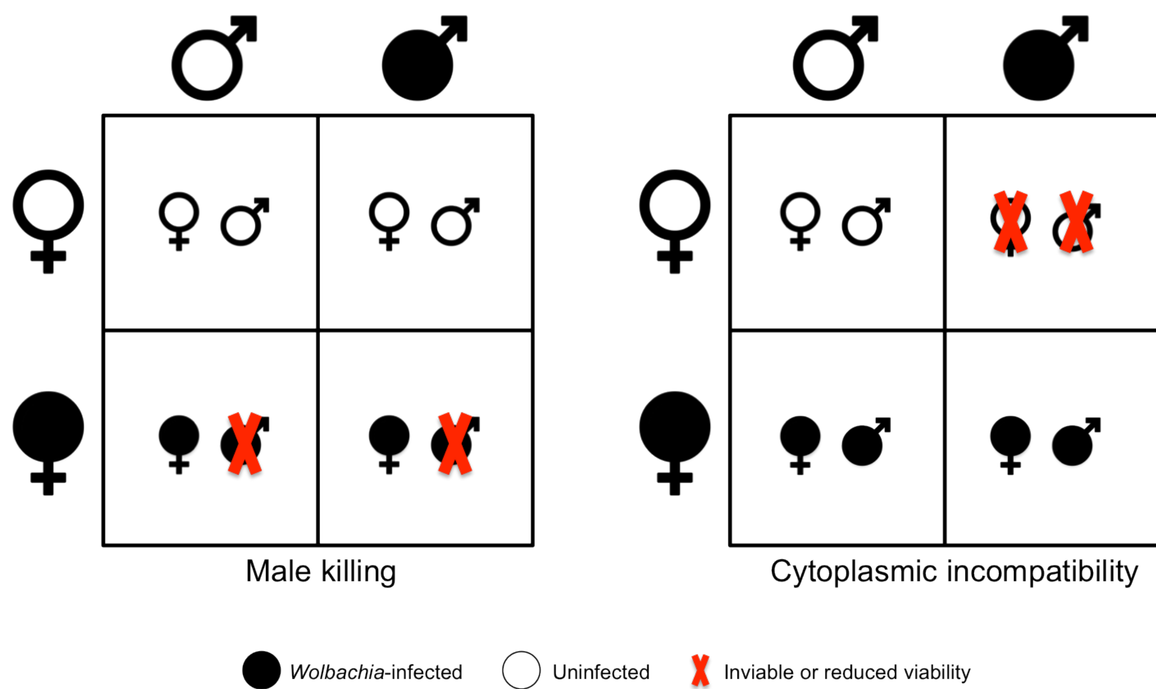


Figure I-2. Two major *Wolbachia*-induced reproductive manipulations. Male killing results in the reduced viability of male offspring from infected females, while cytoplasmic incompatibility causes decreased viability of both sexes following matings between uninfected females and infected males.

Mediators of cytoplasmic incompatibility

Despite the scientific recognition of *Wolbachia*-induced CI for more than 40 years (Yen and Barr, 1971) and a basic understanding of the cytological defects that unfold in affected embryos (Serbus et al., 2008), the bacterial effectors that cause CI are completely unknown. Theoretical models explaining CI patterns abound (Bossan et al., 2011, Poinsot et al., 2003), and a number of candidate effectors have been postulated, but no definitive evidence has been uncovered even though more than a dozen genes have been empirically tested (Yamada et al., 2011). Nonetheless, a substantial body of genomic, transcriptomic, proteomic, and bioinformatic data has been amassed, enabling a relatively unbiased approach to selecting candidate effectors. To identify possible effectors of CI, we used these datasets to narrow the list of potential genes from over 1000 to only 14, before using predicted gene functions to select six candidates for functional characterization. Since *Wolbachia* has never been successfully transformed, we used a transgenic *Drosophila melanogaster* system to express these candidates in fruit flies and test their ability to cause or rescue CI. Ongoing progress in these experiments is described in Chapter V.

Wolbachia evolution during host switching

One major area of interest in *Wolbachia* research is the use of this microbe as a method of biocontrol against insect vectors that transmit pathogens to humans and livestock (Iturbe-Ormaetxe et al., 2011). The success of this strategy depends on using CI to drive the spread of desirable *Wolbachia* strains into wild populations of mosquitos.

Interestingly, *Wolbachia* does not naturally infect the most important mosquito vectors for human illnesses such as dengue fever, chikungunya virus, filariasis, and malaria. However, when artificially infected, *Wolbachia* can block disease transmission by either shortening mosquito lifespan, thereby preventing life cycle progression of the pathogens (McMeniman et al., 2009), or by inhibiting pathogen development directly through unclear mechanisms (Kambris et al., 2009, Moreira et al., 2009). Although horizontal transmission of *Wolbachia* between insect species is common on evolutionary timescales (Vavre et al., 1999), artificial and stable infection of mosquitoes has been difficult, only succeeding in some species after first infecting mosquito cell lines and passaging these cells for years (McMeniman et al., 2009, McMeniman et al., 2008). This suggests that host switching in *Wolbachia* may be associated with a genetic adaptation that makes the bacterium more fit in its new host, something that must evolve over time in culture. To investigate this adaptation, we sequenced the genome of *Wolbachia* strain *w*Mel in its native host, *D. melanogaster*, and after eight years of culture in an *Aedes aegypti* cell line. The genetic changes associated with this adaptation and their functional consequences are investigated in Chapter VI.

Conclusions and future directions

The research described here makes significant inroads into the understanding of *Wolbachia* biology, but many questions remain. Chapter VII will describe several potential avenues for future investigation, as well as communicate preliminary data for some of these studies.

CHAPTER II. THE COMPLEXITY OF VIRUS SYSTEMS: THE CASE OF ENDOSYMBIONTS*

Abstract

Host-microbe symbioses involving bacterial endosymbionts comprise some of the most intimate and long-lasting interactions on the planet. While restricted gene flow might be expected due to their intracellular lifestyle, many endosymbionts, especially those that switch hosts, are rampant with mobile DNA and bacteriophages. One endosymbiont, *Wolbachia pipientis*, infects a vast number of arthropod and nematode species and often has a significant portion of its genome dedicated to prophage sequences of a virus called WO. This phage has challenged fundamental theories of bacteriophage and endosymbiont evolution, namely the phage Modular Theory and bacterial genome stability in obligate intracellular species. WO has also opened up exciting windows into the tripartite interactions between viruses, bacteria, and eukaryotes.

Introduction

Bacterial endosymbionts that replicate within eukaryotic cells are extremely widespread in nature. In addition to the endosymbiont-derived organelles of mitochondria and chloroplasts, more recently-evolved bacterial endosymbionts are abundant in nature, occurring in virtually all eukaryotic hosts (Taylor et al., 2011).

---

* This chapter is published in *Curr Opin Microbiol*, 2012, 15(4): 546-52. Seth R. Bordenstein contributed to the authorship of this chapter.

Historically, obligate intracellular endosymbionts were thought to be devoid of mobile and laterally acquired DNA given their isolated niche, but recent studies have shown that the ecology of bacterial endosymbionts significantly influences the amount of their genome populated by mobile elements such as phages (Figure II-1) (Newton and Bordenstein, 2011, Bordenstein and Reznikoff, 2005). Here, we discuss the prevalence of endosymbiont viruses and focus on recent reports describing the evolution, host interactions, and scientific applications of one of the most widespread and well-studied endosymbiont viruses, phage WO.

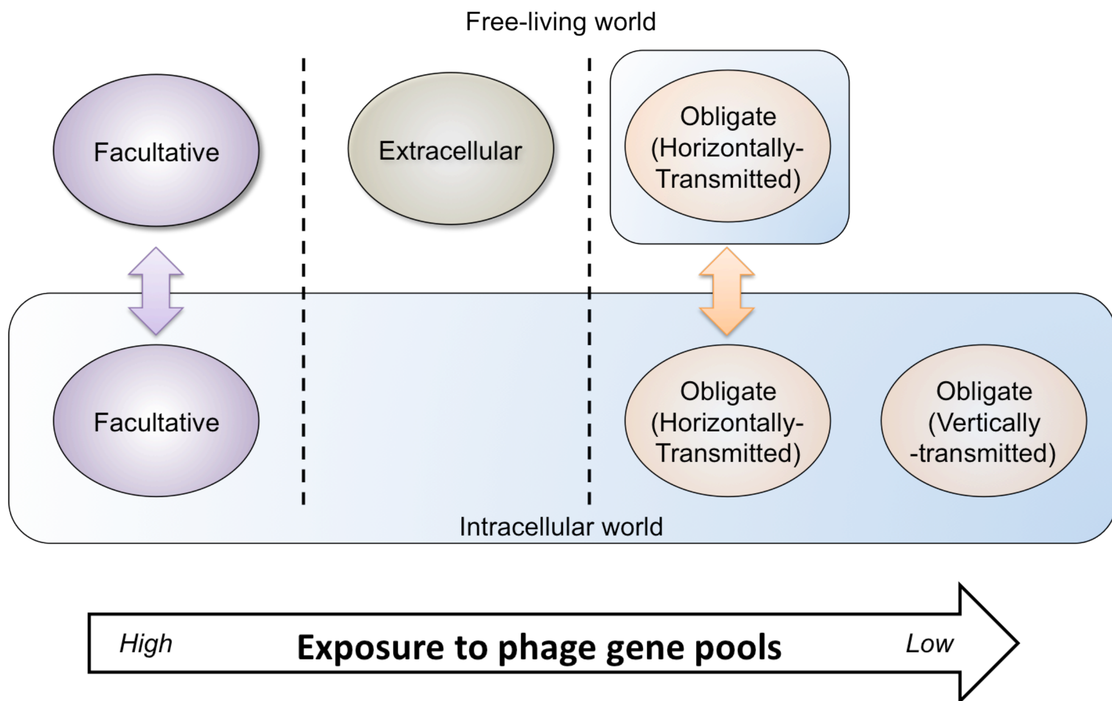


Figure II-1. Effects of microbial ecology on exposure to phage gene pools. Facultative intracellular bacteria have the largest exposure to bacteriophage genes due to their flexible lifestyle involving both the free-living and intracellular environments; thus, they have the greatest amount of mobile DNA in their genomes. Extracellular bacteria have an intermediate amount of mobile DNA, while obligate intracellular bacteria have the least. However, intracellular bacteria that switch hosts and can be horizontally transmitted often retain a large quantity of mobile DNA including phages.

Prevalence of phages in endosymbionts

Bacteriophages are the most abundant biological entity on Earth, outnumbering their unicellular hosts by at least an order of magnitude (Clokie et al., 2011). Although free-living bacteria are less restrictive targets for phages, the most recent survey of mobile genetic elements in bacteria has shown that many endosymbionts possess equal amounts of mobile DNA including phages (Newton and Bordenstein, 2011). While endosymbionts that are strictly vertically transmitted from mother to offspring, such as *Buchnera*, *Wigglesworthia*, and *Blochmannia*, often lack phages, the genomes of those that switch hosts, such as *Chlamydia*, *Rickettsia*, *Phytoplasma*, and *Wolbachia*, often contain a high percentage of mobile DNA (Bordenstein and Reznikoff, 2005). Indeed, 21% of the genome of the *w*Pip strain of *Wolbachia pipientis* is comprised of mobile DNA, including five prophages (Klasson et al., 2008), and phages are present in *Chlamydia pneumoniae* isolates throughout the globe (Rupp et al., 2007). Additionally, endosymbionts not currently infected by phages often show evidence of past infections. For example, *w*Bm, the *Wolbachia* strain infecting the nematode *Brugia malayi*, has at least six phage pseudogenes even though it currently lacks a whole prophage (Kent and Bordenstein, 2010, Foster et al., 2005). Even mitochondria, which have been obligate endosymbionts for over a billion years, possess genes that likely were derived from ancient bacteriophages (Shutt and Gray, 2006).

The phages of *Wolbachia* in particular merit closer examination for several reasons: (1) *Wolbachia* is likely the most widespread endosymbiotic genus on the planet, infecting an estimated 66% of all arthropod species (Hilgenboecker et al., 2008) as well as most medically and agriculturally important nematodes (Bandi et al., 2001). (2) Many

Wolbachia strains are rampantly infected with a group of temperate dsDNA bacteriophages named WO (Kent and Bordenstein, 2010, Gavotte et al., 2007). (3) *Wolbachia* exhibit numerous influences on their hosts that ensure their spread as reproductive parasites (Werren et al., 2008) (see section below on reproductive parasitism), and WO may play a role in these effects (Pichon et al., 2012). (4) WO phages have several potential applications as tools for understanding endosymbiont evolution and manipulating their biology.

Evolution of WO

The availability of a large number of sequenced WO phages and *Wolbachia* genomes has enabled a close examination of WO genome structure and evolution (Kent et al., 2011a). There are five strains of *Wolbachia* in which active phage particle production has been demonstrated (Bordenstein et al., 2006, Fujii et al., 2004, Sanogo and Dobson, 2006, Gavotte et al., 2007), each of which contains prophages with complete head, baseplate, and tail gene modules essential for proper phage function (Figure II-2). Interestingly, *Wolbachia* strains that harbor a complete WO phage usually have additional WO prophages that are degenerate, transcriptionally inactive (Biliske et al., 2011), and, with a few exceptions (Klasson et al., 2009b, Klasson et al., 2008), not closely related to other prophages in the same strain (Kent et al., 2011a).

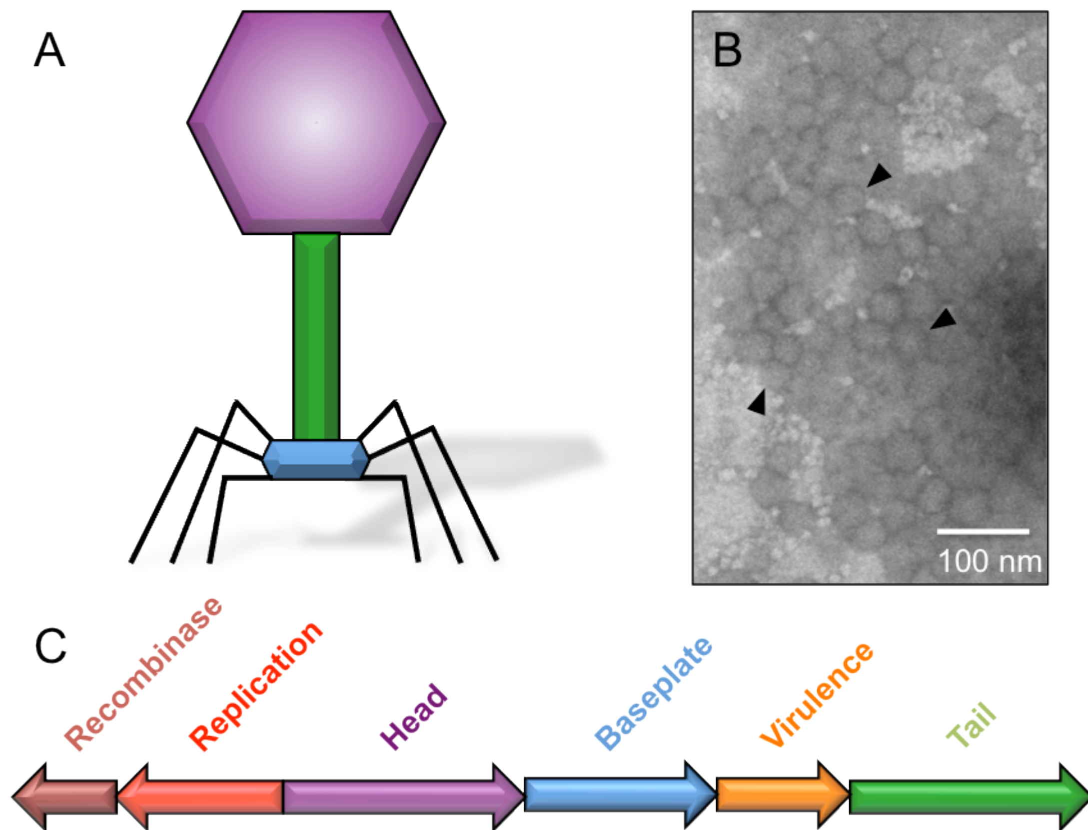


Figure II-2. WO particle and genome structure. (A) Typical appearance of a tailed bacteriophage, color-coded by structural groups. (B) Electron micrograph of WO particles. Examples of phage particles are indicated with arrowheads. Shown is WO isolated from *w*CauB in the moth Ephestia kuehniella. Photo courtesy of Sarah Bordenstein. (C) The modular genome of phage WO. Relative portions of the genome dedicated to individual modules and the modules' orientation and arrangement are shown for WOCauB2. Other WO strains have modules in differing arrangements and orientations and some may lack various modules all together. Not all genes are shown.

It is commonly understood that dsDNA bacteriophages evolve mainly through frequent horizontal gene transfer of contiguous sets of unrelated genes with a similar function (i.e. tail genes, head genes, lysis genes, etc) between phages in a common gene pool. This tenet is termed the Modular Theory (Botstein, 1980). However, analysis of 16 WO sequences revealed for the first time that, although WO phages are modular, they do not evolve according to the Modular Theory but rather through point mutation, intragenic

recombination, deletion, and purifying selection (Figure II-3) (Kent et al., 2011a). Thus, although WO is prevalent in *Wolbachia*, its obligate intracellular niche limits the exposure of WO to other phages with which to recombine. Indeed, all evolutionarily recent horizontal transfer events among WO phages are between co-infections of intracellular bacteria in the same eukaryotic host, reflecting the fact that endosymbionts have relatively little interaction with free-living bacteria or their phages (Figure II-3). Examples of these transfers include a 52 kb phage transfer between *Wolbachia* strains *w*VitA and *w*VitB coinfecting the parasitic wasp *Nasonia vitripennis* (Kent et al., 2011b), and multiple phage transfers between coinfecting *Wolbachia* strains in natural populations of the leaf beetle *Neochlamisus bebbianae* (Chafee et al., 2010). Transfer can also occur between different species of obligate or facultative intracellular bacteria, such as between *Wolbachia* and a plasmid from a *Rickettsia* endosymbiont of the tick *Ixodes scapularis* (Figure II-4) (Ishmael et al., 2009).

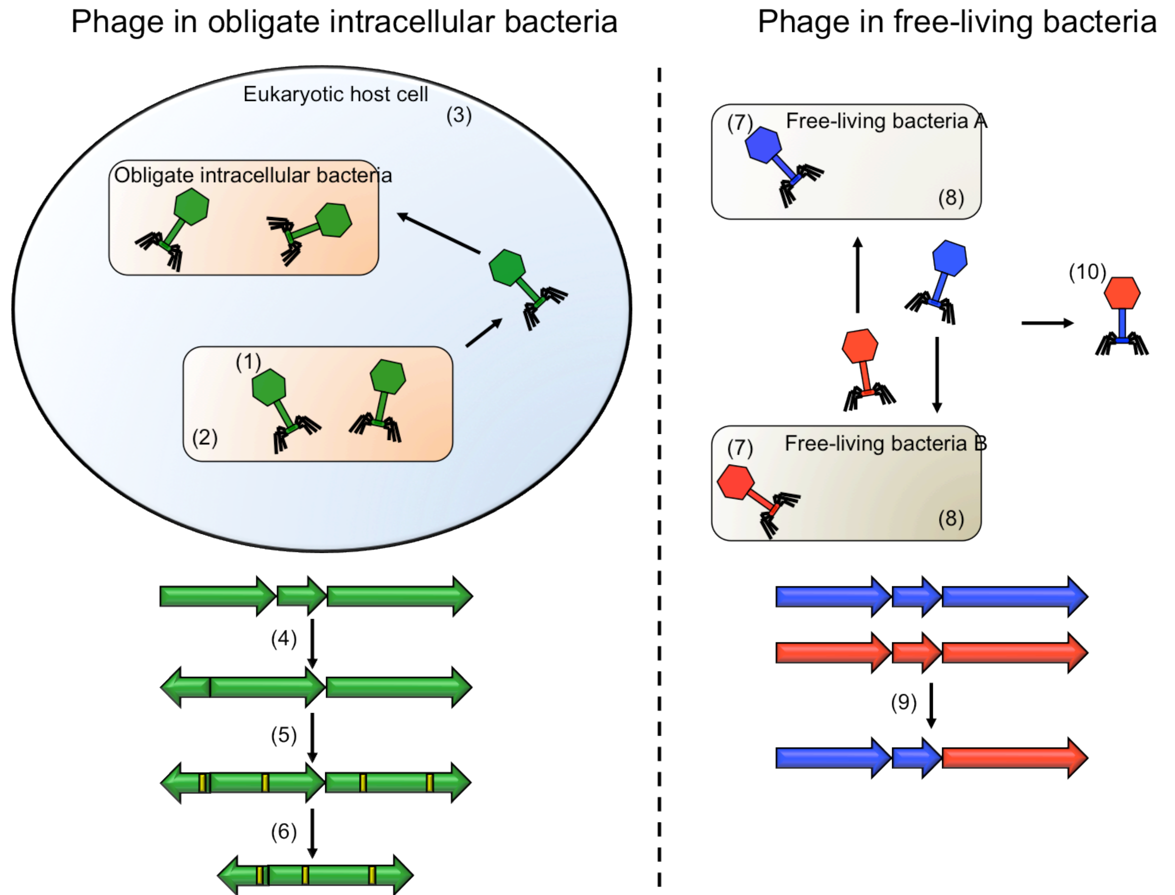


Figure II-3. Evolution of bacteriophages in endosymbionts and free-living bacteria. Bacteriophages (1) of endosymbionts (2) are restricted in their interactions with other phages due to the barrier of the eukaryotic host membrane (3). Their genomes evolve mainly through recombination (4), point mutation (5), and deletion (6). Bacteriophages (7) of free-living bacteria (8) can more freely interact with each other facilitating modular gene exchange (9) and forming viruses consisting of parts of each parent strain (10). Thus, free-living but not endosymbiont phages evolve by the Modular Theory.

In addition to transfer of phages between bacteria, lateral gene transfer of *Wolbachia* genes into their eukaryotic hosts' genomes is surprisingly common, with *Wolbachia* genes found in at least seven insect species and four nematode species (Dunning Hotopp et al., 2007, Klasson et al., 2009a, Fenn et al., 2006, Nikoh et al., 2008). These inserts range in size from less than 500bp in *Nasonia* to nearly the entire

Wolbachia genome in *Drosophila ananassae* (Dunning Hotopp et al., 2007). Interestingly, these transfers often include WO prophage regions (Dunning Hotopp et al., 2007) or sequences adjacent to WO in the *Wolbachia* genome (Figure II-4) (Klasson et al., 2009a). Given the extensive host range of these endosymbionts, many more as yet undiscovered horizontal transfer events are likely.

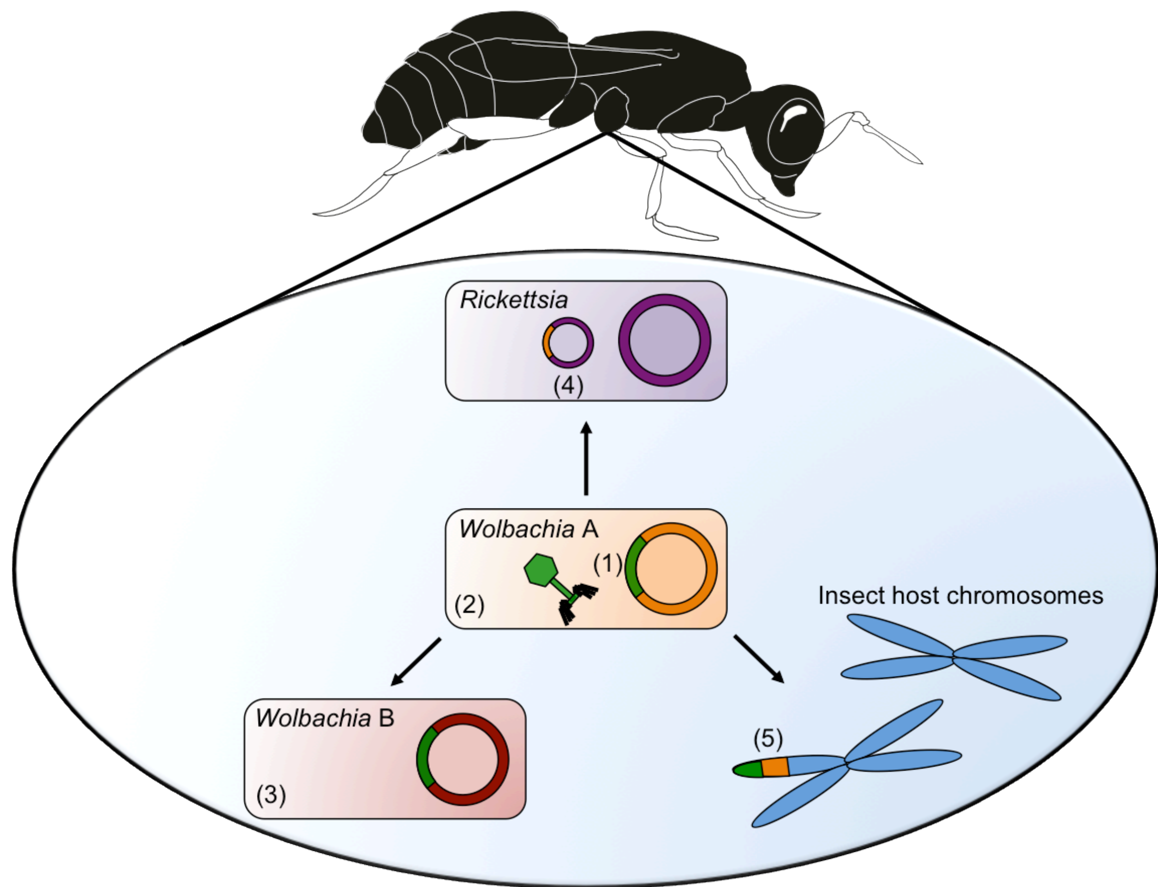


Figure II-4. Examples of gene flow between WO, *Wolbachia*, and insects. WO prophage sequences (1) have been transferred between coinfections of different *Wolbachia* strains (2 and 3) on several occasions. Additionally, *Wolbachia* genes have been transferred to a *Rickettsia* plasmid (4), and both WO and *Wolbachia* genes have been found in multiple insect host genomes (5).

Involvement of WO in reproductive parasitism

Perhaps the most tantalizing concept in the study of WO is the idea that WO may influence the biology of not only *Wolbachia*, but also *Wolbachia*'s arthropod hosts. *Wolbachia* have evolved several mechanisms for manipulating their hosts' reproduction to ensure their spread and maintenance in a population by increasing the evolutionary fitness of *Wolbachia*-transmitting females (Werren et al., 2008). These mechanisms include (1) male killing (male offspring die during embryogenesis), (2) feminization (genetic males develop into fertile females), (3) parthenogenesis (virgin females produce all female broods) and (4) cytoplasmic incompatibility (CI), an asymmetrical crossing incompatibility in which offspring of *Wolbachia*-infected males and uninfected females die during early embryogenesis. The idea that WO could be involved in these manipulations is based on the precedent that bacteriophages commonly encode virulence factors and other genes promoting the fitness of both phage and its host (Boyd and Brussow, 2002). Even endosymbiont phages may provide such a function. For example, APSE, a phage of *Hamiltonella defensa*, defends *H. defensa*'s host, the aphid *Aphidius ervi*, against parasitic wasps, likely through a phage-encoded toxin of unknown mechanism (Oliver et al., 2009, Degnan and Moran, 2008). Additionally, *Wolbachia* genomes and especially WO prophage regions are replete with ankyrin-repeat proteins (Iturbe-Ormaetxe et al., 2005), a motif known to mediate diverse protein-protein interactions in eukaryotes (Al-Khodor et al., 2010); thus they could facilitate *Wolbachia*'s reproductive manipulation of its hosts.

Wolbachia-induced reproductive manipulations are remarkably complex. For example, bidirectional CI blocks the production of offspring between two insects

harboring different strains of *Wolbachia* in some cases but not others (Zabalou et al., 2008), leading to several theories for how CI functions. The Lock and Key Model postulates that numerous combinations of modification (mod) factors alter arthropod sperm such that they cannot develop in uninfected eggs, while rescue (resc) factors repair this defect if the egg is infected with a compatible strain of *Wolbachia* (Zabalou et al., 2008, Poinsot et al., 2003). Another theory, the Goalkeeper Model, posits that only two factors exist, but that their concentration or activity level accounts for incompatibility between some strains (Bossan et al., 2011). In any case, these intricate CI patterns have enabled a search for correlations between strain compatibility and WO, although the results have been somewhat contradictory (Saridaki et al., 2011, Sanogo and Dobson, 2006, Sanogo et al., 2005, Gavotte et al., 2007).

One hypothesis is that a WO DNA methyltransferase gene may encode the mod and/or resc factors of CI (Saridaki et al., 2011). This theory fits well with the fact that sperm DNA appears to be modified in the hosts of mod+ *Wolbachia* strains and that DNA methylation is altered during feminization of the leafhopper species *Zyginidia pullula* when infected with *Wolbachia*, although methylation patterns have not yet been investigated in CI (Negri et al., 2009). Remarkably, all resc+ group A *Wolbachia* examined have a WO-encoded *met2* methyltransferase gene, whereas resc- *Wolbachia* do not. However, this correlation does not extend to group B *Wolbachia*, suggesting that if *met2* is the resc factor in group A, it is not universal or its equivalent in group B has not yet been recognized (Saridaki et al., 2011). The *met2* gene has been constitutively expressed in *Drosophila melanogaster* and was unable to cause or rescue CI in *w*Mel-infected flies (Yamada et al., 2011). Nevertheless, there are several additional genes

found in mod+, resc+ strains but not mod-, resc- strains (Iturbe-Ormaetxe et al., 2005), so it remains possible that the WO methyltransferase is involved in CI but requires additional proteins. Examination of transcription of WO genes has shown differential expression of haplotypes of a capsid gene, *orf7*, between sexes, strains, and life stages of *Culex pipiens* mosquitoes (Sanogo and Dobson, 2006); however, there has been no obvious correlation between *orf7* haplotypes and CI patterns in several species (Sanogo et al., 2005, Gavotte et al., 2007). Perhaps most damning to the hypothesis that WO underlies reproductive parasitism is the fact that some *Wolbachia* strains without WO still manipulate host reproduction (Gavotte et al., 2007). Therefore, if WO genes are directly involved in arthropod reproductive manipulation, their effect is likely not universal in all strains, but could be part of a larger interplay with other *Wolbachia* genes and host factors.

Even if WO genes are not directly involved in reproductive manipulations, there is significant evidence that WO indirectly influences CI by controlling *Wolbachia* densities in the host, a theory termed the Phage Density Model (Kent and Bordenstein, 2010). In *w*VitA, which infects *Nasonia vitripennis* and contains active, lytic WO, densities of *Wolbachia* and WO are inversely related, as are *Wolbachia* densities and CI severity (Bordenstein et al., 2006). Interestingly, altering *Wolbachia* environmental factors does not abolish this three-way interaction. Introgression to move the wVitA strain from its native host into a related species of wasp, N. giraulti increased *Wolbachia* load, decreased WO densities, and increased CI (Chafee et al., 2011), while rearing insects at temperature extremes had the opposite effect (Bordenstein and Bordenstein, 2011). In wPip-infected *Culex pipiens* mosquitoes under conditions where WO is not

lytic, this correlation is not seen (Walker et al., 2009). These results strongly suggest that lytic WO influences CI by altering *Wolbachia* densities. Additionally, this interaction is influenced by host factors in a tripartite relationship between WO, *Wolbachia*, and their insect host.

Applications of WO

One of the greatest limitations in *Wolbachia* research is the inability to successfully transform these bacteria. Until the *Wolbachia* genome can be manipulated, it is unlikely that fundamental questions regarding the mechanism of CI and other aspects of *Wolbachia* biology will be definitively answered. Fortunately, WO offers hope as an avenue for accomplishing this genetic manipulation. Recombinases and attachment sites for WO integration have been identified that could be exploited to this end (Tanaka et al., 2009), although there is significant diversity in recombinases and no integration site common to all WO prophages (Kent et al., 2011a). The large size of the WO genome, diversity of phage sequences, and intracellular lifestyle of *Wolbachia* are all obstacles to overcome, but development of a WO DNA-delivery vector would be a colossal advance in the study of *Wolbachia*.

WO also has a potential therapeutic application. Although *Wolbachia* is a reproductive parasite in most arthropods, in many parasitic nematodes, including those causing filariasis and river blindness in humans, *Wolbachia* is mutualistic and required for the nematodes' reproduction (Werren et al., 2008). Indeed, elimination of *Wolbachia* with antibiotic therapies has been successful in treating filarial diseases (Taylor et al.,

2010). WO may encode useful gene products for inhibiting *Wolbachia*, as phages often express numerous proteins for manipulation and inhibition of their hosts. Potential candidates in WO include lysozymes, which lyse bacterial cell walls (Fischetti, 2010), and patatins, which have a phospholipase activity (Nevalainen et al., 2008). Lysozymes have been identified in two WO phages, while patatins are nearly universal in WO (Kent et al., 2011a). An understanding of how WO manipulates and lyses *Wolbachia* may enable development of small molecules with similar functions, or the use of WO's own proteins as therapeutics if they can be accompanied by an appropriate delivery system (Borysowski and Gorski, 2010).

Conclusions

Given the abundance and range of *Wolbachia* and its phage WO, a firm grasp of the biology in this system will be important for understanding endosymbiont viruses in general and their interactions with their hosts. WO has already tested fundamental questions in evolutionary theory and hinted at fascinating host interactions at multiple levels of symbiotic relationships. Further study of WO and perhaps use of WO as a tool for genetic manipulation will no doubt lead to even more intriguing discoveries in the future.

CHAPTER III. PARALLEL HORIZONTAL TRANSFER AND FUNCTION OF AN ANTIBACTERIAL GENE ACROSS THE TREE OF LIFE[†]

Abstract

Though horizontal gene transfer (HGT) is widespread, genes and taxa experience biased rates of transferability. Curiously, independent transmission of the same DNA to archaea, bacteria, eukaryotes, and viruses is extremely rare and often defies ecological and functional explanations. Here, we demonstrate that a bacterial lysozyme family integrated independently in all domains of life across diverse environments, generating the only glycosyl hydrolase 25 muramidases in plants and archaea. In *in vitro* cultivated archaea from hydrothermal vents, muramidase transcription is upregulated in coculture with bacteria and recombinant lysozyme exhibits broad-spectrum antibacterial action in a dose-dependent manner. Similar to bacterial transfer of antibiotic resistance genes, transfer of a potent antibacterial gene across the universal tree seemingly bestows a niche-transcending adaptation that trumps the barriers against parallel HGT to all domains. The discoveries also comprise the first characterization of an antibacterial gene in archaea and support the pursuit of antibiotics in this underexplored group.

[†] Lisa J. Funkhouser-Jones, Kristen Brileya, Anna-Louise Reysenbach, and Seth R. Bordenstein contributed to the authorship of this chapter.

Introduction

Horizontal gene transfer (HGT) is rampant among prokaryotes and phages and is an important mechanism for acquisition of new genes and functions (Popa and Dagan, 2011), including the shuttling of antibiotics and antibiotic resistance between bacteria (Clardy et al., 2009). Instances of interdomain horizontal transfer of diverse genes between two domains of life or between viruses and their hosts are also increasingly documented (Nelson et al., 1999, Husnik et al., 2013, Dunning Hotopp et al., 2007, Wu et al., 2013, Gladyshev et al., 2008, Bratke and McLysaght, 2008, Danchin et al., 2010). A few of these transfers are functionally characterized, but the biological activity, selective advantages, and ecological contexts of many interdomain HGT events remain poorly characterized (Dunning Hotopp, 2011, Keeling and Palmer, 2008). In comparison to these intradomain or interdomain highways of HGT (Beiko et al., 2005), independent transmission of the same gene family to archaea, bacteria, eukaryotes, and viruses is extremely uncommon and subject to apparently rare events throughout the history of life (Moran et al., 2012, Lundin et al., 2010, Koonin et al., 2003, McClure, 2001, McDonald et al., 2012). When taken together, genome-enabled studies suggest that horizontal gene transfers (HGTs) are biased and experience a frequency gradient that decreases from within domain > between two domains > between all domains of life (Bruto et al., 2013, Zhaxybayeva and Doolittle, 2011, Puigbo et al., 2009, Andam and Gogarten, 2013, Andam and Gogarten, 2011).

One significant question then is why do highways of intra- or interdomain transfers occur more frequently than transfers to all domains in the universal tree of life? There are at least two explanations. First, recurrent transfer of the same gene family may

be limited by incompatible mechanics of gene transfer (e.g., transduction, transfection, plasmid exchange) between domains compared to within domains. However, the individual success of gene transfers between any two domains of life, e.g. archaea and bacteria (Nelson et al., 1999, van Wolferen et al., 2013), bacteria and eukaryote (Andersson, 2005, Bordenstein, 2007, Gladyshev et al., 2008, Danchin et al., 2010), and archaea and eukaryote (Schonknecht et al., 2013, Andersson et al., 2003), suggests that this barrier may be minimal. Second, the selective barriers against HGT of the same gene to multiple taxa and preservation of the gene through evolutionary time are multifaceted given the potential costs associated with HGT (Baltrus, 2013) and that each recipient may not benefit from the trait conferred. Thus, there may be very few niche-transcending genes (Wiedenbeck and Cohan, 2011), defined as genes that are useful in different physiological capabilities, cellular structures, and ecological niches that repeatedly increase fitness of each recipient across the whole diversity of life.

Among the few putative cases, there is a pore-forming toxin domain that appears to have been anciently transferred between diverse lineages (Moran et al., 2012). However, the distribution of the transfer across the tree of life is unclear because archaea sequences were not included in phylogenetic analyses due to low phylogenetic support values. Other candidate genes encode proteins involved in nucleotide metabolism, intramembrane proteolysis, or membrane transport, but the transfer events defy clear interpretations due to their deep antiquity in evolutionary time and the confounding issues of ancient paralogy (Lundin et al., 2010, Koonin et al., 2003, McClure, 2001, McDonald et al., 2012). Moreover, these transfers are often not functionally validated in the recipient taxa.

Here we demonstrate for the first time, to our knowledge, that a functional antibacterial gene family scattered across the tree of life in diverse ecological contexts. This bacterial gene encodes a glycosyl hydrolase 25 (GH25) muramidase, a peptidoglycan-degrading lysozyme that hydrolyzes the 1,4-β-linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine in the bacterial cell wall. Typically found in bacteria (Cantarel et al., 2009), the lytic enzyme classically functions in cell division and cell wall remodeling (Vollmer et al., 2008), while in bacteriophages they lyse bacterial peptidoglycan at the end of the phage life cycle (Fastrez, 1996). Although members of the GH25 muramidase family have been noted in other taxa (Korczynska et al., 2010, Nikoh et al., 2010), extensive analysis of their evolutionary history and functions have not been undertaken. We hypothesized that, similar to the transfer of antibiotic resistance genes between bacteria, the transfer of antibacterial genes from bacteria to archaea and to eukaryotes bestows a niche-transcending adaptation that trumps the resilient barriers against independent HGTs of the same type of gene across the tree of life.

Results and Discussion

A Bacterial GH25 Muramidase Is Present in All Domains of Life

During a homology search, we uncovered 75 nonredundant homologs (E-values ≤ 10^{-12}) of a bacterial GH25 muramidase in disparate taxa across the tree of life, indicating possible HGT of a bacterial gene to both eukaryotic and archaeal species as well as to

phages. Putative HGT events were identified in the genomes of the plant *Selaginella moellendorffii* (Banks et al., 2011), the deep-sea hydrothermal vent archaeon *Aciduliprofundum boonei* (Reysenbach et al., 2006), the pea aphid *Acyrthosiphon pisum* (International Aphid Genomics Consortium, 2010, Nikoh et al., 2010), and several species of fungi such as *Aspergillus oryzae* (Machida et al., 2005). We verified the presence of the lysozyme gene in natural populations of selected HGT recipients by PCR and sequencing of the GH25 muramidase domain (Figure III-1), including *Aciduliprofundum* field samples harvested from hydrothermal vents worldwide. We detected lysozyme genes in 9 out of 12 field isolates of *Aciduliprofundum* from deep-sea vents in the Atlantic and Pacific oceans, 5 out of 6 species in the plant genus *Selaginella*, and 8 out of 9 aphid species in the subfamily Aphidinae (Figure III-1,

Table III-1). Sequenced field isolate lysozymes were relatively similar to each other in each clade, with 74% pairwise identity (385aa alignment) amongst *Aciduliprofundum* sequences, 85.1% (203aa alignment) amongst three intact *Selaginella* sequences, and 87.3% identity (93aa alignment) amongst Aphidinae sequences. We also found lysozymes in two additional WO phages as part of an ongoing next generation sequencing project of *Wolbachia* viruses (unpublished data).

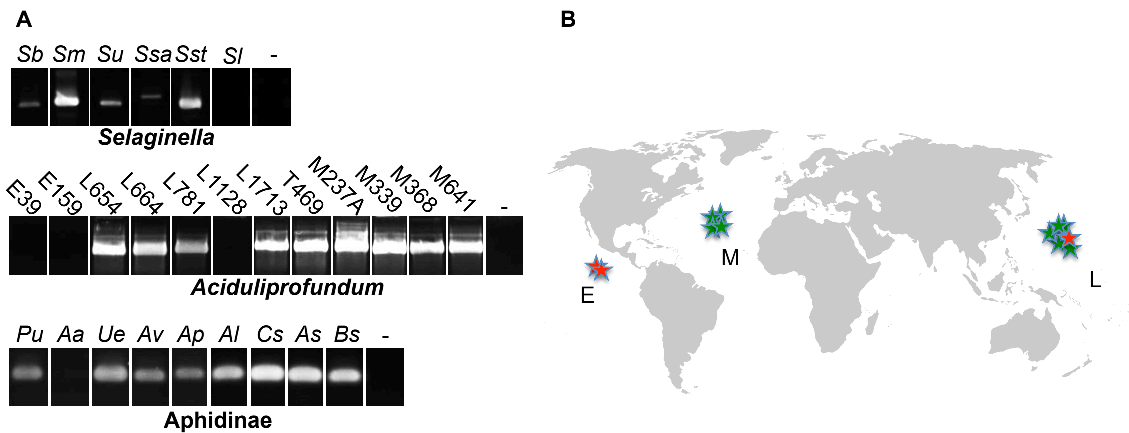


Figure III-1. Presence of HGT lysozyme genes in field samples. (A) PCR amplifications of portions of the GH25 muramidase domain in the indicated taxa. All amplifications were Sanger sequenced to confirm integration. Primers used are listed in Table III-2. Abbreviations: *Sb*: *S. braunii*, *Sm*: *S. moellendorffii*, *Su*: *S. uncinata*, *Ssa*: *S. sanguinolenta*, *Sst*: *S. stauntoniana*, *Sl*: *S. lepidophylla*, E: East Pacific Rise, L: Lao Spreading Center, M: Mid-Atlantic Ridge, *Pu*: *Pleotrichophorus utensis*, *Aa*: *Artemisaphis artemisicola*, *Ue*: *Uroleucon erigeronensis*, *Av*: *Aphis varians*, *Ap*: *Acyrthosiphon pisum*, *Al*: *Aphis lupini*, *Cs*: *Cedoaphis sp.*, *As*: *Aphthargelia symphoricarpi*, *Bs*: *Braggia sp.*, - denotes water only control. **(B)** World map with approximate locations of *A. boonei* field samples. Those that tested positive for the GH25 muramidase domain are indicated by green stars and those that tested negative are indicated by red stars. Map is a public domain image from Wikimedia Commons.

Table III-1. Field samples tested for presence of lysozyme gene. *Aciduliprofundum* strains were cultured from samples obtained from hydrothermal vents in the indicated regions. *Selaginella* species were obtained from a commercial nursery (Plant Delights Nursery, North Carolina, USA). Aphid samples were generously provided by Patrick Abbot (Vanderbilt University, Nashville, TN). WO sequences were pulled from high throughput phage sequencing preparations of laboratory *Wolbachia* strains from the indicated insect species (unpublished data).

Taxon	Isolate, strain, or species	Origin/Distribution
Aciduliprofundum	Lau09-654	Eastern Lau Spreading Center deep-sea vents
	Lau09-664	
	Lau09-781	
	Lau09-1713	
	A. boonei-T469	
	Lau09-1128	
	Mar08-237A	Mid-Atlantic Ridge deep-sea vents
	Mar08-339	
	Mar08-368	
	Mar08-641	
	Epr07-39	East Pacific Rise deep-sea vents
	Epr07-159	
Selaginella	*S. moellendorffii*	China
	S. braunii	China
	S. uncinata	China
	S. lepidophylla	North America
	S. sanguinolenta	Japan
	S. stauntoniana	China
Aphidinae	*Acyrthosiphon pisum*	Worldwide
	Pleotrichophorus utensis	United States
	Artemisaphis artemisicola	North America
	Uroleucon erigeronensis	North America
	Aphis varians	North America
	Aphis lupini	United States
	Cedoaphis sp.	North America
	Aphthargelia symphoricarpi	North America
	Braggia sp.	United States
WO	WORiA	*Drosophila simulans*
	WOCauB3	*Cadra cautella*
	WOVitA4	*Nasonia vitripennis*

Table III-2. Primers used in this study.

Target	Forward Primer (5'-3')	Reverse Primer (5'-3')
S. moellendorffii GH25 muramidase	ATGGACGTAAGTAGCT ACCAAGG	TCAGCCTTTGGCGAGCT TC
Aciduliprofundum GH25 muramidase, degenerate	ATGTKTCCCACTGGCA GG	CCACCCTGTCATCGTAG AAGA
Aphid GH25 muramidase, degenerate	CTYTGGGGAGCATAYC ATTTTGG	TTTTWCCATCKGTRTAY TGCCATAA
Aciduliprofundum GH25 muramidase integration	GGGTGCCTCTCCTCCAA TCCCC	CCACTCACCCCCGATAC ATTTCC
S. moellendorffii GH25 muramidase integration	ATGGCGTTTCATTGCTT GATCTTT	GTTGTAACATTTTTGCGC TGGAGTA
A. boonei GH25 muramidase, qPCR	TCCCACTGGCAGGGAA ATGTGAACT	ATCCTGATGCGTGTGCC TTCTCCA
A. boonei elongation factor 1α, qPCR	TGTTCATCGGCCATGTT GACCACG	GCTCTTTCCGAGTTTCTC TGCCTCCT

To rule out spurious bacterial contamination in these genomes and to confirm genomic integration of the lysozyme gene, we employed direct sequencing of PCR products amplified using primers inside the lysozyme gene paired with primers outside the gene for *Aciduliprofundum* field samples and *S. moellendorffii*. Incorporation of the lysozyme gene was verified in all cases tested (Figure III-2). Additionally, *Aciduliprofundum* samples were grown in strict monocultures as determined by 16S amplicon monitoring. Integration of the *A. pisum* lysozyme has been previously established (Nikoh et al., 2010). Flanking genes in the recipient genomes were non-bacterial on either side of the transferred lysozyme in each case (Figure III-3), with two exceptions. A bordering gene in *A. boonei*, ADP-ribose-1"-monophosphatase (App-1), possesses both bacterial and archaea homologs and a phylogenetic analysis suggests HGT unrelated to the lysozyme transfer (Figure III-4A). This transfer was likely between Archaea and Thermotogae bacteria. The second exception is a GH2 hydrolase gene adjacent to the lysozyme in *A.*

oryzae. This hydrolase has bacterial homologs in the phylum Actinobacteria, and recapitulates the same phylogenetic pattern as seen in the GH25 muramidase (see below, Figure III-4B). Thus, it is likely that the lysozyme and GH2 hydrolase were transferred to fungi in a single event.

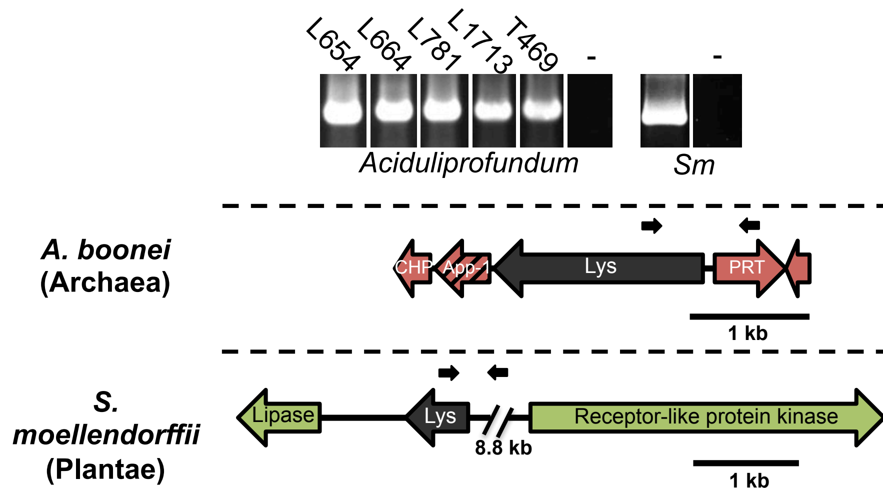


Figure III-2. PCR amplifications testing genomic integration with primers within and outside of lysozyme genes. Primers used are listed in Table III-2 and binding sites are indicated in gene diagrams with small black arrows. All integrations were confirmed with Sanger sequencing. Abbreviations: *Sm*: *S. moellendorffii*, L: Lao Spreading Center, - denotes water only control, CHP = conserved hypothetical protein, App-1 = ADP-ribose-1”monophosphatase, PRT = phosphoribosyltransferase.

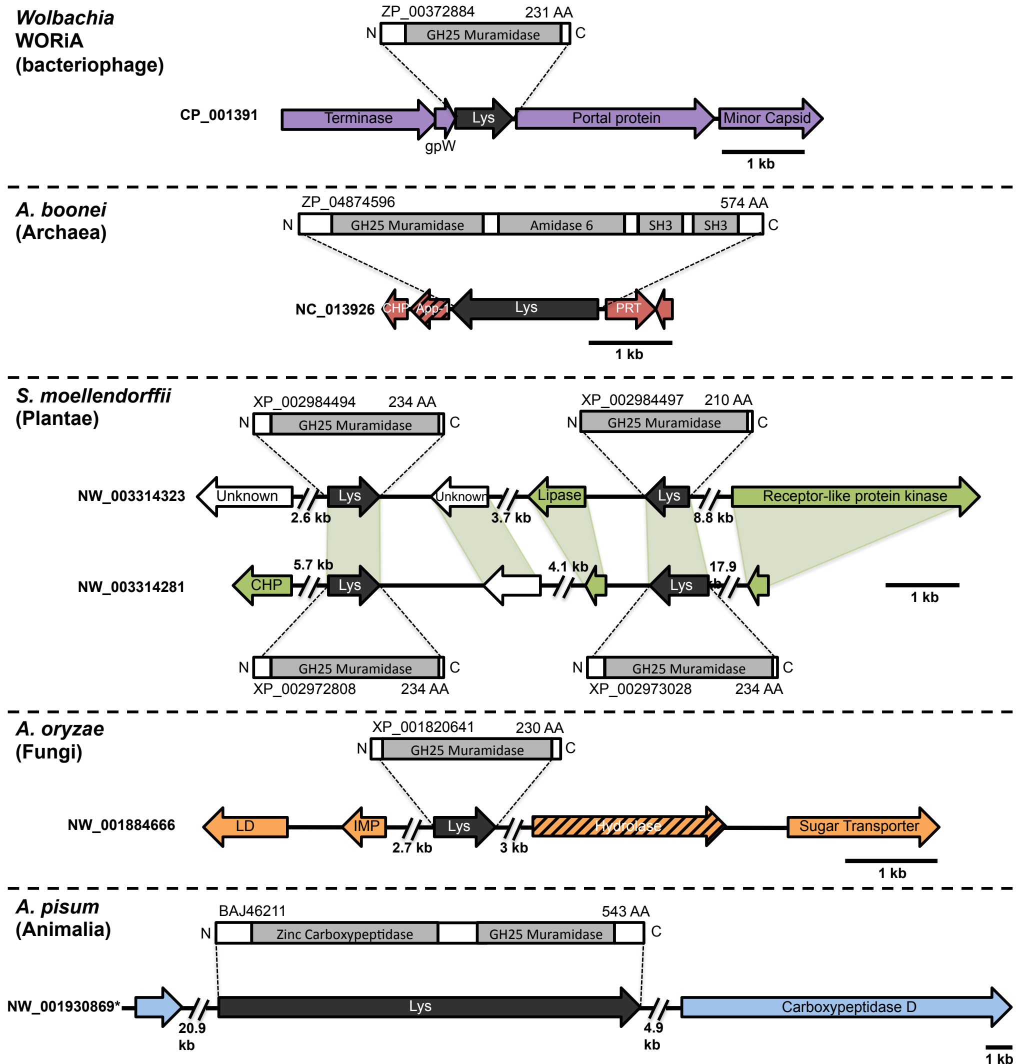


Figure III-3. Architecture of HGT candidates and surrounding genes. Each arrow represents an open reading frame transcribed from either the plus strand (arrow pointing right) or the minus strand (arrow pointing left). The color of the arrow indicates the taxa the gene is found in based on its closest homologs. Black = Eubacteria, purple = virus, red = Archaea, green = Plantae, Orange = Fungi, Blue = Insecta, white = no known homologs, dashed line = present in multiple domains. The length of the arrows and intergenic regions are drawn to scale except where indicated with broken lines. The four paralogs of the lysozyme in *S. moellendorffii* occur on two genomic scaffolds with light green bands connecting homologous genes. Abbreviations: Lys: lysozyme, gpW = phage baseplate assembly protein W, SH3: Src homology domain 3, App-1 = ADP-ribose-1”-monophosphatase, PRT = phosphoribosyltransferase, LD = leucoanthocyanidin dioxygenase; IMP = integral membrane protein. A protein diagram for each lysozyme is drawn to scale with the light gray regions highlighting a conserved protein domain. **A. pisum* diagram is based on Acyr_1.0 assembly and transcription data (Nikoh et al., 2010); the annotation in Acyr_2.0 is different.

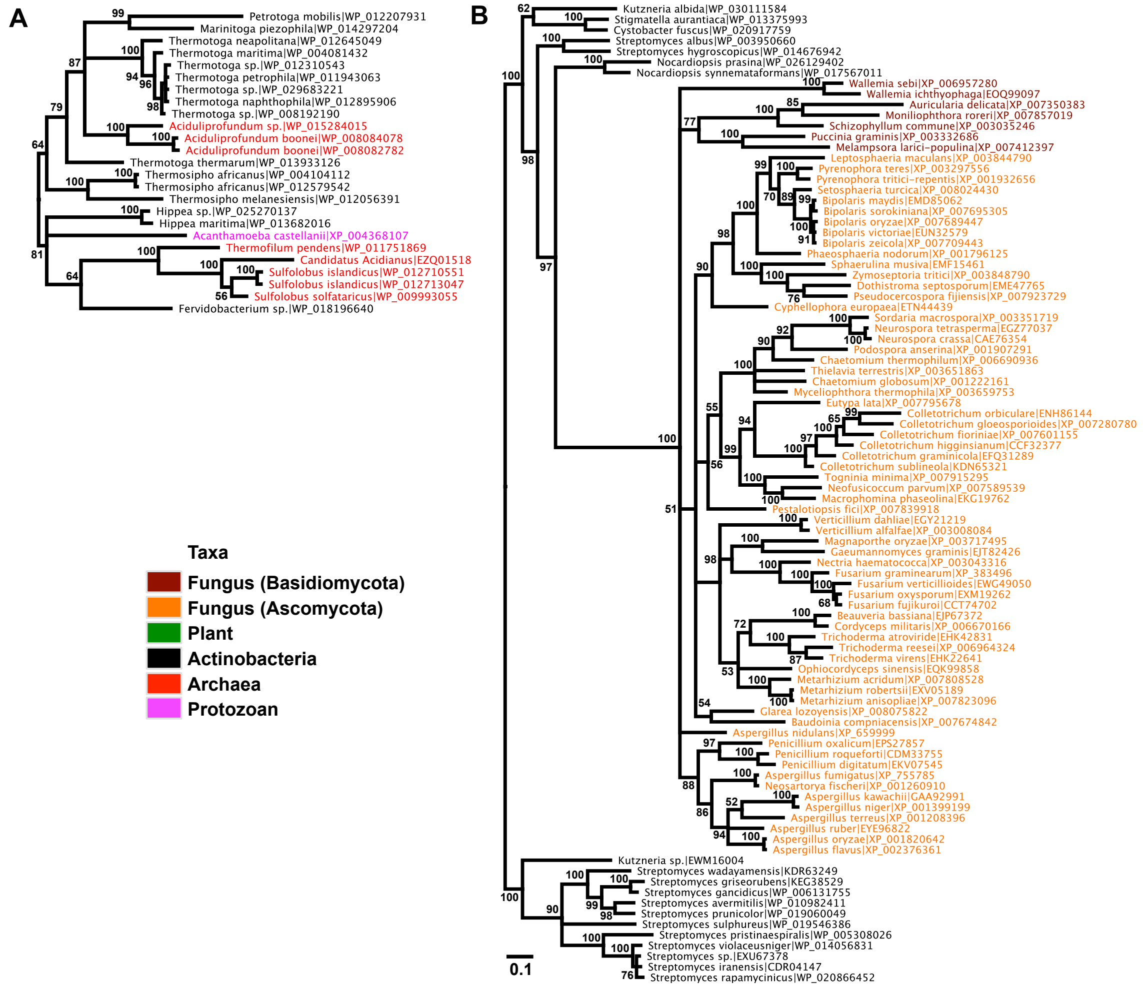


Figure III-4. Protein phylogeny of neighboring genes to transferred lysozymes. (A) App-1 phylogeny based on alignment of 141aa without indels consisting of top E-value hits to blastp using *A. boonei* App-1 as the query. Taxon of origin for each amino acid sequence is indicated by color. Posterior probability is indicated at all nodes with values above 50. Branch lengths represent number of substitutions per site as indicated by scale bar. Tree is arbitrarily rooted. **(B)** GH2 hydrolase phylogeny based on an alignment of 188aa without indels consisting of top E-value hits to blastp using *A. oryzae* hydrolase as the query.

Non-bacterial GH25 Muramidases Arose From HGT

To establish parallel HGT, i.e., the independent transfer of the same gene family to multiple lineages, we conducted a phylogenetic analysis on 86 GH25 muramidase sequences using Bayesian and maximum likelihood inference methods (Figure III-5A). We combined non-redundant *Aciduliprofundum*, *Selaginella*, and WO sequences obtained from PCR and Sanger sequencing with blastp results to reconstruct the phylogeny. Additionally, transferred lysozymes in nonbacterial taxa were used as queries to identify homologs and make a second set of phylogenetic trees to confirm the HGT (Figure III-5B-D). Three key results emerge from these phylogenetic analyses: (i) at least three independent instances of interdomain HGT of the bacterial GH25 muramidase occurred in nonbacterial taxa (*Aciduliprofundum*, *Selaginella*, and *Insecta*) as well as a number of transfers to bacteriophages, (ii) vertical transmission of the transferred gene ensues in some descendant taxa (i.e., *Aciduliprofundum* and *Selaginella*), and (iii) frequent HGT of the muramidase between bacterial clades accompanies the interdomain transfer, indicating that transfer across the tree of life is the norm for this niche-transcending gene family.

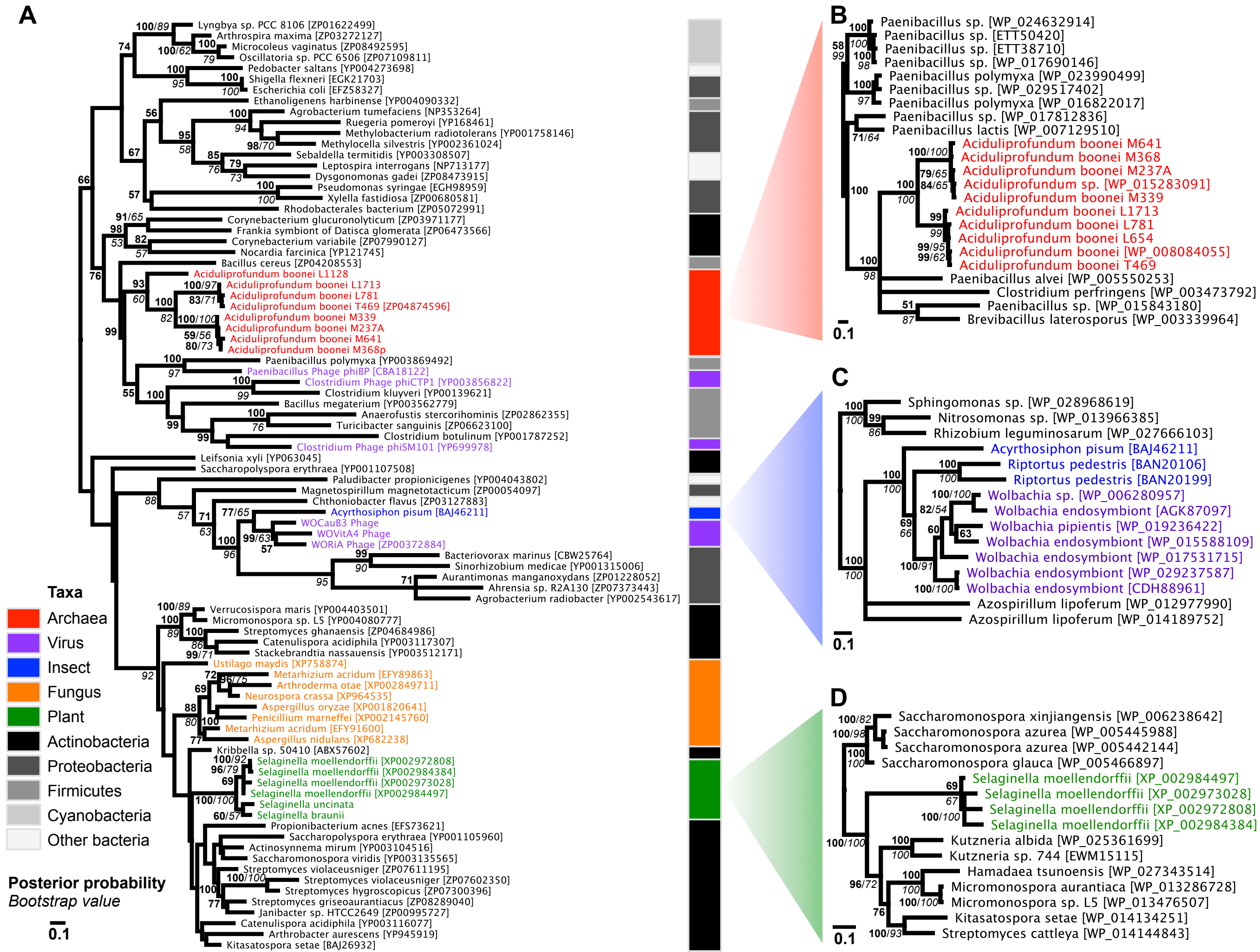

A
Lyngbya sp. PCC 8106 [ZP01622499]
Arthrospira maxima [ZP03272127]
Microcoleus vaginatus [ZP08492595]
Oscillatoria sp. PCC 6506 [ZP07109811]
Pedobacter saltans [YP004273698]
Shigella flexneri [EGK21703]
Escherichia coli [EFZ58327]
Ethanoligenens harbinense [YP004090332]
Agrobacterium tumefaciens [NP353264]
Ruegeria pomeroyi [YP168461]
Methylobacterium radiotolerans [YP001758146]
Methylocella silvestris [YP002361024]
Sebaldella termitidis [YP003308507]
Leptospira interrogans [NP713177]
Dysgonomonas gadei [ZP08473915]
Pseudomonas syringae [EGH98959]
Xylella fastidiosa [ZP00680581]
Rhodobacterales bacterium [ZP05072991]
Corynebacterium glucuronolyticum [ZP03971177]
Frankia symbiont of Datisca glomerata [ZP06473566]
Corynebacterium variabile [ZP07990127]
Nocardia farcinica [YP121745]
Bacillus cereus [ZP04208553]
Aciduliprofundum boonei L1128
Aciduliprofundum boonei L1713
Aciduliprofundum boonei L781
Aciduliprofundum boonei T469 [ZP04874596]
Aciduliprofundum boonei M339
Aciduliprofundum boonei M237A
Aciduliprofundum boonei M641
Aciduliprofundum boonei M368p
Paenibacillus polymyxa [YP003869492]
Paenibacillus Phage phiBP [CBA18122]
Clostridium Phage phiCTP1 [YP003856822]
Clostridium kluyveri [YP00139621]
Bacillus megaterium [YP003562779]
Anaerofustis stercorihominis [ZP02862355]
Turicibacter sanguinis [ZP06623100]
Clostridium botulinum [YP001787252]
Clostridium Phage phiSM101 [YP699978]
Leifsonia xyli [YP063045]
Saccharopolyspora erythraea [YP001107508]
Paludibacter propionicigenes [YP004043802]
Magnetospirillum magnetotacticum [ZP00054097]
Chthoniobacter flavus [ZP03127883]
Acyrthosiphon pisum [BAJ46211]
WOCauB3 Phage
WOVitA4 Phage
WORiA Phage [ZP00372884]
Bacteriovorax marinus [CBW25764]
Sinorhizobium medicae [YP001315006]
Aurantimonas manganoxydans [ZP01228052]
Ahrensia sp. R2A130 [ZP07373443]
Agrobacterium radiobacter [YP002543617]
Verrucosispora maris [YP004403501]
Micromonospora sp. L5 [YP004080777]
Streptomyces ghanaensis [ZP04684986]
Catenulispora acidiphila [YP003117307]
Stackebrandtia nassauensis [YP003512171]
Ustilago maydis [XP758874]
Metarhizium acridum [EFY89863]
Arthroderma otae [XP002849711]
Neurospora crassa [XP964535]
Aspergillus oryzae [XP001820641]
Penicillium marneffei [XP002145760]
Metarhizium acridum [EFY91600]
Aspergillus nidulans [XP682238]
Kribbella sp. 50410 [ABX57602]
Selaginella moellendorffii [XP002972808]
Selaginella moellendorffii [XP002984384]
Selaginella moellendorffii [XP002973028]
Selaginella moellendorffii [XP002984497]
Selaginella uncinata
Selaginella braunii
Propionibacterium acnes [EFS73621]
Saccharopolyspora erythraea [YP001105960]
Actinosynnema mirum [YP003104516]
Saccharomonospora viridis [YP003135565]
Streptomyces violaceusniger [ZP07611195]
Streptomyces violaceusniger [ZP07602350]
Streptomyces hygroscopicus [ZP07300396]
Streptomyces griseoaurantiacus [ZP08289040]
Janibacter sp. HTCC2649 [ZP00995727]
Catenulispora acidiphila [YP003116077]
Arthrobacter aurescens [YP945919]
Kitasatospora setae [BAJ26932]
Taxa
Archaea
Virus
Insect
Fungus
Plant
Actinobacteria
Proteobacteria
Firmicutes
Cyanobacteria
Other bacteria
Posterior probability
Bootstrap value
0.1
B
Paenibacillus sp. [WP_024632914]
Paenibacillus sp. [ETT50420]
Paenibacillus sp. [ETT38710]
Paenibacillus sp. [WP_017690146]
Paenibacillus polymyxa [WP_023990499]
Paenibacillus sp. [WP_029517402]
Paenibacillus polymyxa [WP_016822017]
Paenibacillus sp. [WP_017812836]
Paenibacillus lactis [WP_007129510]
Aciduliprofundum boonei M641
Aciduliprofundum boonei M368
Aciduliprofundum boonei M237A
Aciduliprofundum sp. [WP_015283091]
Aciduliprofundum boonei M339
Aciduliprofundum boonei L1713
Aciduliprofundum boonei L781
Aciduliprofundum boonei L654
Aciduliprofundum boonei [WP_008084055]
Aciduliprofundum boonei T469
Paenibacillus alvei [WP_005550253]
Clostridium perfringens [WP_003473792]
Paenibacillus sp. [WP_015843180]
Brevibacillus laterosporus [WP_003339964]
0.1
C
Sphingomonas sp. [WP_028968619]
Nitrosomonas sp. [WP_013966385]
Rhizobium leguminosarum [WP_027666103]
Acyrthosiphon pisum [BAJ46211]
Riptortus pedestris [BAN20106]
Riptortus pedestris [BAN20199]
Wolbachia sp. [WP_006280957]
Wolbachia endosymbiont [AGK87097]
Wolbachia pipientis [WP_019236422]
Wolbachia endosymbiont [WP_015588109]
Wolbachia endosymbiont [WP_017531715]
Wolbachia endosymbiont [WP_029237587]
Wolbachia endosymbiont [CDH88961]
Azospirillum lipoferum [WP_012977990]
Azospirillum lipoferum [WP_014189752]
0.1
D
Saccharomonospora xinjiangensis [WP_006238642]
Saccharomonospora azurea [WP_005445988]
Saccharomonospora azurea [WP_005442144]
Saccharomonospora glauca [WP_005466897]
Selaginella moellendorffii [XP_002984497]
Selaginella moellendorffii [XP_002973028]
Selaginella moellendorffii [XP_002972808]
Selaginella moellendorffii [XP_002984384]
Kutzneria albida [WP_025361699]
Kutzneria sp. 744 [EWM15115]
Hamadaea tsunoensis [WP_027343514]
Micromonospora aurantiaca [WP_013286728]
Micromonospora sp. L5 [WP_013476507]
Kitasatospora setae [WP_014134251]
Streptomyces cattleya [WP_014144843]
0.1

Figure III-5. Phylogeny of GH25 muramidase. **(A)** Phylogeny based on alignment of 113aa without indels consisting of top E-value hits to blastp using WORiA phage lysozyme as a query. Taxon of origin for each amino acid sequence is indicated by color. Posterior probability (Bayesian phylogeny) and bootstrap values (maximum likelihood phylogeny) are indicated at all nodes with values above 50. Branch lengths represent number of substitutions per site as indicated by scale bar. Tree is arbitrarily rooted. Iterative phylogenies based on top E-value blastp hits to *A. boonei* lysozyme **(B)**, *A. pisum* lysozyme **(C)**, and *S. moellendorffii* lysozyme **(D)** are also shown.

To statistically validate parallel HGT across the tree of life, we performed a Shimodaira-Hasegawa test (SH-test) (Shimodaira and Hasegawa, 1999) by comparing our consensus tree (Figure III-5) against a hypothetical tree with a binary constraint in which bacterial sequences are monophyletic and separate from monophyletic nonbacterial sequences. In this hypothetical tree consistent with the tree of life, lineage relationships with the bacterial and nonbacterial groups were permissively set as unconstrained. Results indicate that the hypothetical tree is significantly worse than the HGT tree, as expected ($P < 0.01$, D(LH) = -133.9, SD = 31.5). We repeated this analysis with the hypothetical tree compared to 100 bootstrapped HGT trees and found the hypothetical tree was also worse than each of these trees ($P < 0.01$). Thus, the null hypothesis of vertical descent is rejected under the most permissive conditions.

We observed that each interdomain HGT event (Figure III-5) occurred between taxa that coexist in the same ecological niche, a likely prerequisite for HGT. For instance, the *A. boonei* lysozyme is in a clade dominated by Firmicutes whose members can be common in deep ocean sediments (Orcutt et al., 2011a). The *A. pisum* lysozyme clade includes *Wolbachia* prophages and Proteobacteria, which are common endosymbionts of aphids and other insects (Augustinos et al., 2011, Gomez-Valero et al., 2004, Wang et al.,

2014). The *S. moellendorffii* plant lysozyme is closely related to Actinobacteria, which are dominant microbes in soil (Bulgarelli et al., 2013). These associations, while not proof of HGT, establish interactions that may have facilitated the transfers. Although the phylogenetic pattern of the GH25 muramidase found in fungi is consistent with HGT (Figure III-5A, Figure III-6), the transfer occurred anciently in fungal evolution prior to the divergence of Dikarya, as the domain is present in both Basidiomycota and Ascomycota, but not other fungal phyla. As is the case with most putative ancient transfers, the deep branches of the tree are poorly supported and a definitive donor taxon cannot be established. However, a nucleotide-level phylogeny also supports HGT from an ancestral Actinobacterium (Figure III-7).

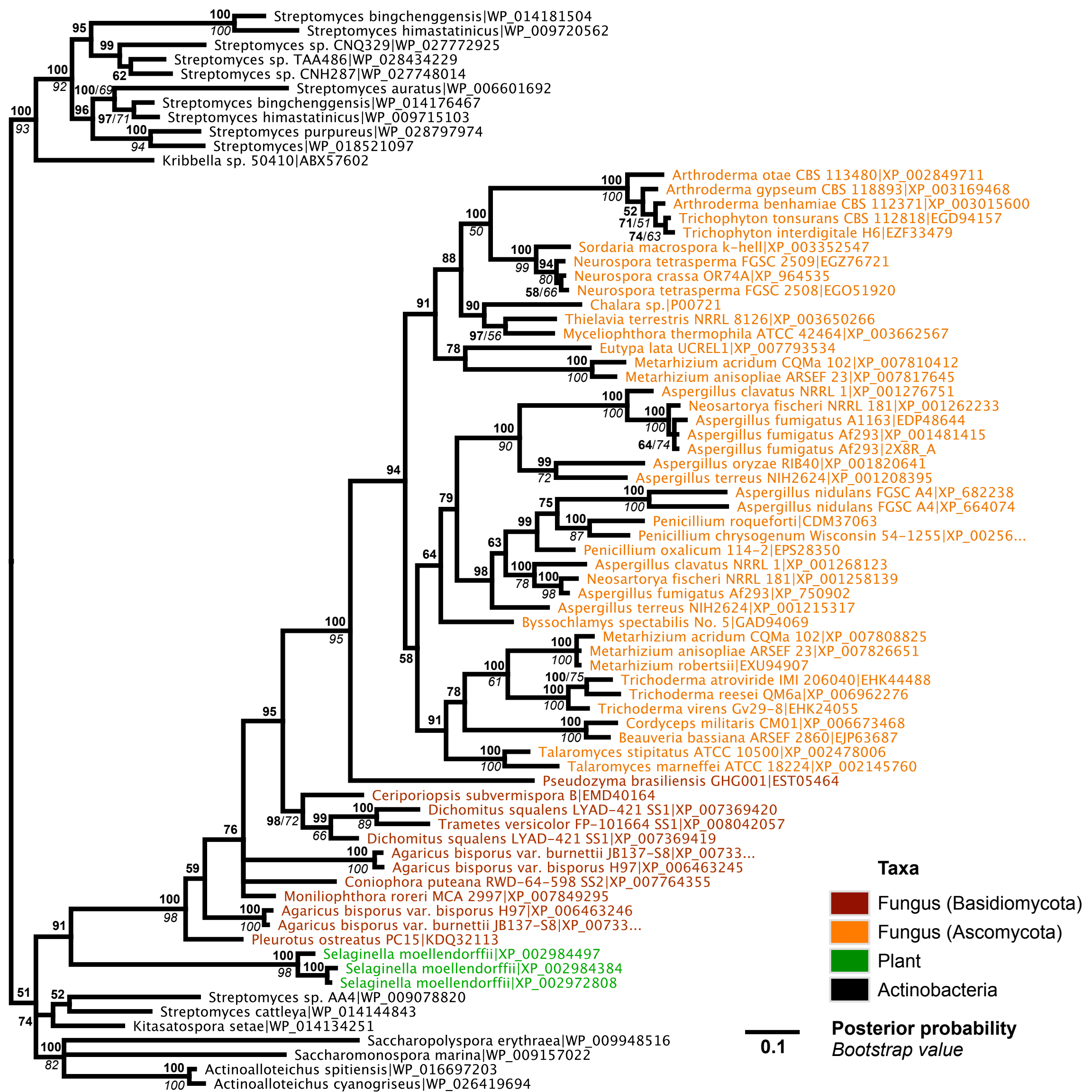


Figure III-6. Protein phylogeny of *A. oryzae* GH25 muramidase and relatives. Phylogeny based on alignment of 186aa without indels consisting of top E-value hits to blastp using *A. oryzae* lysozyme as a query. Taxon of origin for each amino acid sequence is indicated by color. Posterior probability (Bayesian phylogeny) and bootstrap values (maximum likelihood phylogeny) are indicated at all nodes with values above 50. Branch lengths represent number of substitutions per site as indicated by scale bar. Tree is arbitrarily rooted.

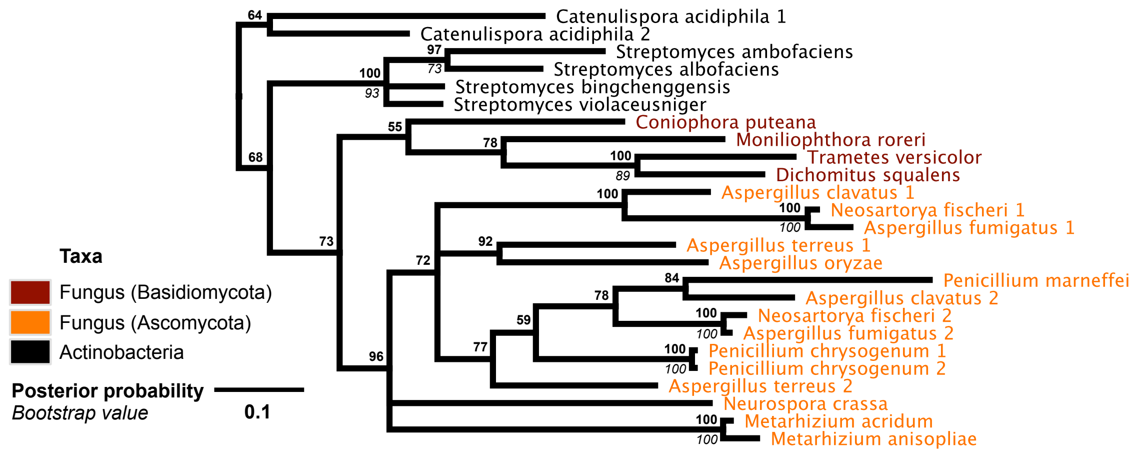


Figure III-7. DNA phylogeny of *A. oryzae* GH25 muramidase and relatives. Phylogeny based on alignment of 282bp without indels consisting of top E-value hits to blastn using *A. oryzae* lysozyme exon 2 as a query. Taxon of origin for each nucleic acid sequence is indicated by color. Posterior probability (Bayesian phylogeny) and bootstrap values (maximum likelihood phylogeny) are indicated at all nodes with values above 50. Branch lengths represent number of substitutions per site as indicated by scale bar. Tree is arbitrarily rooted.

Interestingly, the lysozyme gene in the aphid *A. pisum* consists of a fusion of a bacterial GH25 muramidase domain and a eukaryotic carboxypeptidase domain. The gene includes five introns (Nikoh et al., 2010), none of which interrupt the GH25 domain, consistent with a relatively recent HGT event and the absence of the gene from most sequenced insects. The lysozyme in the fungus *A. oryzae*, meanwhile, contains only a single intron, but it does interrupt the GH25 domain, consistent with the domain's long association with fungi from the subkingdom Dikarya. We found no evidence of a GH25 muramidase in 323 sequenced archaeal genomes spanning all the major phyla and sister taxa to *A. boonei* (Reysenbach et al., 2006, Flores et al., 2012). This lack of homology does not appear to be due to insufficient representation of Archaea diversity, as the 323 members span all of the major phyla: Crenarchaeota (56), Euryarchaeota (205),

Nanoarchaeota (10), and Thaumarchaeota (39). Indeed, if vertical descent were assumed for a recent Bayesian phylogeny of Archaea with sequenced genomes (Brochier-Armanet et al., 2011), this would require at least 10 independent losses of the lysozyme gene, an assumption that is certainly less parsimonious than a single HGT event. Moreover, the relative divergence of the small subunit rRNA gene in *A. boonei* compared to the putative bacterial HGT donors is greater than the relative divergence of the lysozyme gene (Figure III-8), a finding that is inconsistent with both genes being transmitted by vertical descent. Also, there are no other homologs beyond those presented in this study in 132 plant genomes, and only one insect species with additional homologs out of 109 insect genomes. Thus, if the lysozyme were present in the last common ancestor of all domains, it would require the unlikely loss of the gene in dozens of lineages while maintaining it in an exceedingly small number of species. In summary, the presence of a GH25 muramidase in nonbacterial species represents a series of recurrent, independent horizontal gene transfer events derived from diverse, ecologically associated bacteria.

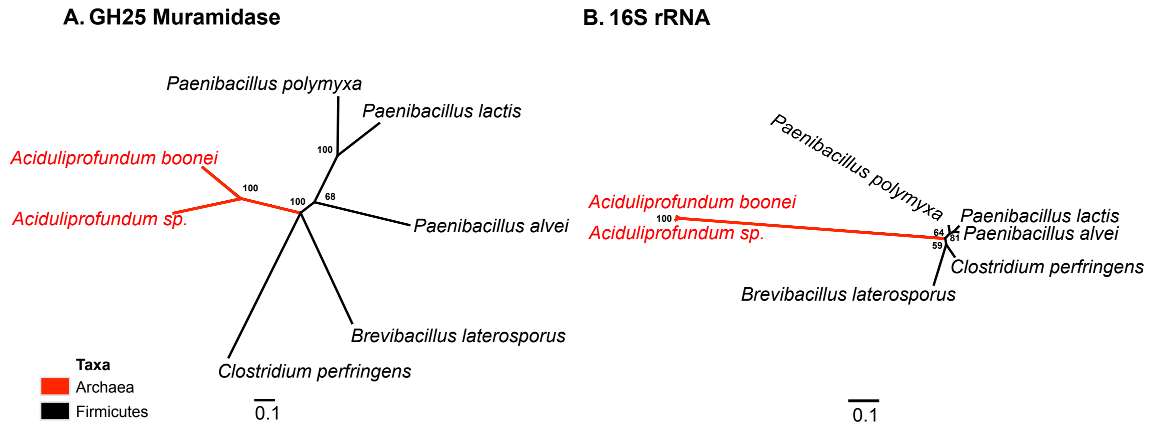


Figure III-8. Comparison of GH25 muramidase and rRNA divergence. (A) Unrooted Bayesian phylogeny of the GH25 muramidase from *A. boonei* and selected relatives, based on an alignment of 185aa without indels. Taxon of origin for each nucleic acid sequence is indicated by color. Posterior probability is indicated at all nodes with values above 50. Branch lengths represent number of substitutions per site as indicated by scale bar. **(B)** Unrooted Bayesian phylogeny of the 16S rRNA gene for the same taxa as in (A), based on an alignment of 1,156bp without indels.

A. boonei GH25 Muramidase Is Antibacterial

We next undertook a series of experiments to test the hypothesis that the transferred muramidase functions as an antibacterial. Since HGT frequently results in pseudogenized and nonfunctional genes (Kondrashov et al., 2006, Nikoh et al., 2010, Nikoh et al., 2008, Dunning Hotopp et al., 2007), we first investigated the amino acid sequences for preserved antibacterial action of the transferred lysozymes in nonbacterial genomes. We aligned all 86 GH25 muramidase sequences to identify conserved sites (Figure III-9A). We then mapped the conserved amino acids to a three-dimensional structure prediction of the *A. boonei* GH25 muramidase domain (Figure III-9B). Highly conserved residues (>85% identity between all taxa) invariably mapped to the previously identified active site pocket (Martinez-Fleites et al., 2009). Conservation was also evident

for structure predictions of other GH25 muramidases in the phylogeny such as *S. moellendorffii* (Figure III-9C).

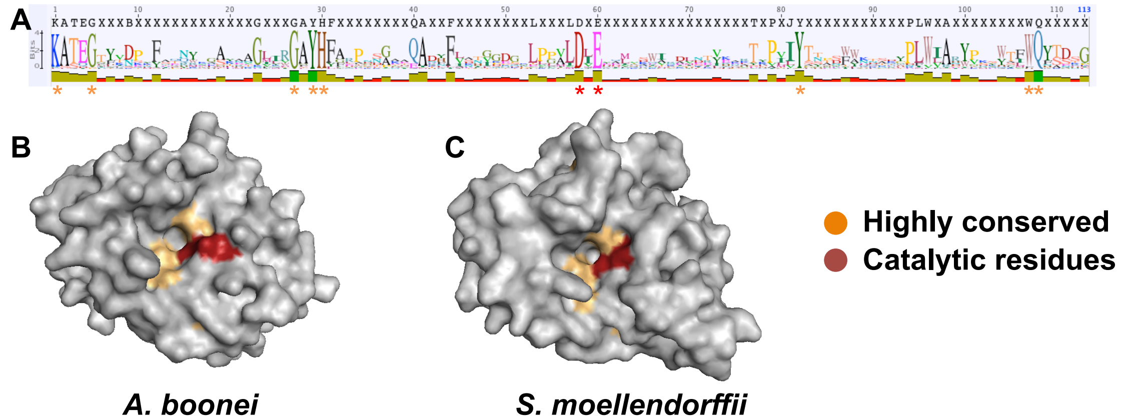


Figure III-9. Conservation of *A. boonei* GH25 muramidase domain. (A) Consensus alignment of 86 GH25 muramidases with insertions and deletions removed. Conservation is indicated by amino acid symbol size and bar graphs below the consensus sequence. Active site residues and highly conserved amino acids modeled below are indicated with red and orange asterisks, respectively. **(B)** Space-filling model of the active site face of the predicted structure of *A. boonei* GH25 muramidase domain and **(C)** *S. moellendorffii* GH25 muramidase domain. Active site residues are indicated in red and the eight additional residues most highly conserved across all 86 proteins are orange.

Second, we cloned, expressed, and purified the GH25 muramidase domain from the archaea *A. boonei* as well as from closely related homologs in *P. polymyxa* and PhiBP. We obtained each muramidase in a pure elution (Figure III-10) and tested for antibacterial action against a range of bacterial species. As predicted, the *A. boonei* GH25 muramidase efficiently killed several species of bacteria in the phylum Firmicutes - the putative donor group of the gene (Figure III-11A). The bacterial inhibition by *A. boonei* GH25 muramidase was more potent than the positive control, chicken egg white

lysozyme, and was dose-dependent (Figure III-11B). Bacterial and phage muramidases did not elicit antibacterial killing, similar to cyan fluorescent protein and buffer-only negative controls. Bacteria typically use a large protein complex to limit their lysozymes' activity to the septum during cell division (Uehara and Bernhardt, 2011), and PhiBP phage has a documented spectrum of activity limited only to a *P. polymyxa* strain unavailable for our analyses (Halgasova et al., 2010). As expected, the *A. boonei* GH25 muramidase did not exhibit antibacterial activity against Gram-negative species or Gram-positive species outside of the families Bacillaceae and Paenibacillaceae, which was equivalent to the killing range of chicken egg white lysozyme with the exception of the Actinobacterium *M. luteus* (Figure III-12).

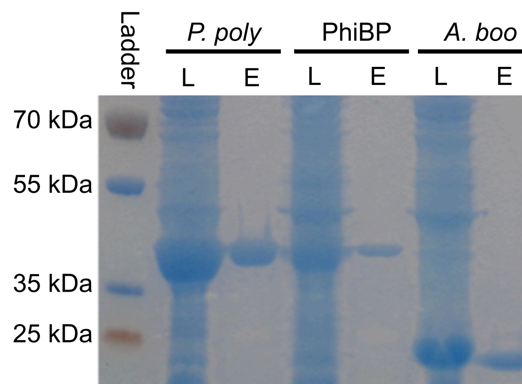


Figure III-10. Lysozyme purifications. PAGE gel stained with GelCode blue before and after purification of 6x-histidine tagged enzymes using nickel affinity chromatography. L = crude *E. coli* lysate expressing the indicated lysozyme, E = elution after lysozyme purification. *P. poly* = *P. polymyxa* lysozyme, PhiBP = bacteriophage PhiBP lysozyme, *A. boo* = *A. boonei* GH25 domain.

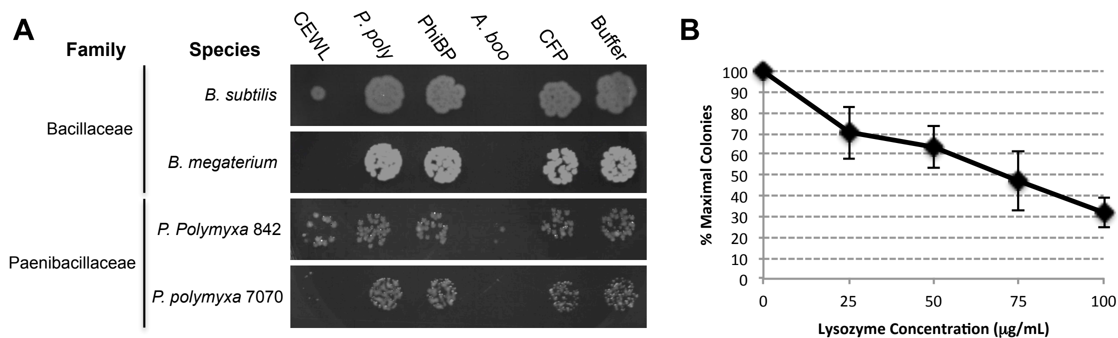


Figure III-11. Antibacterial action of *A. boonei* GH25 muramidase domain against Firmicutes. (A) Bacteria of the specified strain/species incubated overnight on tryptic soy agar after a 20-minute liquid preincubation with the proteins indicated. Genera: *B* = *Bacillus*, *P* = *Paenibacillus*. Proteins: CEWL = chicken egg white lysozyme, *P. poly* = *P. polymyxa* lysozyme, PhiBP = bacteriophage PhiBP lysozyme, *A. boo* = GH25 domain of *A. boonei* lysozyme, CFP = cyan fluorescent protein. Images are representative of at least three independent experiments. **(B)** Dose-dependence of *A. boonei* GH25 muramidase antibacterial action. *B. subtilis* colony survival after incubation with *A. boonei* GH25 muramidase at the indicated concentrations for 20 min at 37 °C. N = 10 for each concentration. P <0.001 for linear model fit. Error bars are +/- SEM. Bacterial strains used are listed in Table III-3.

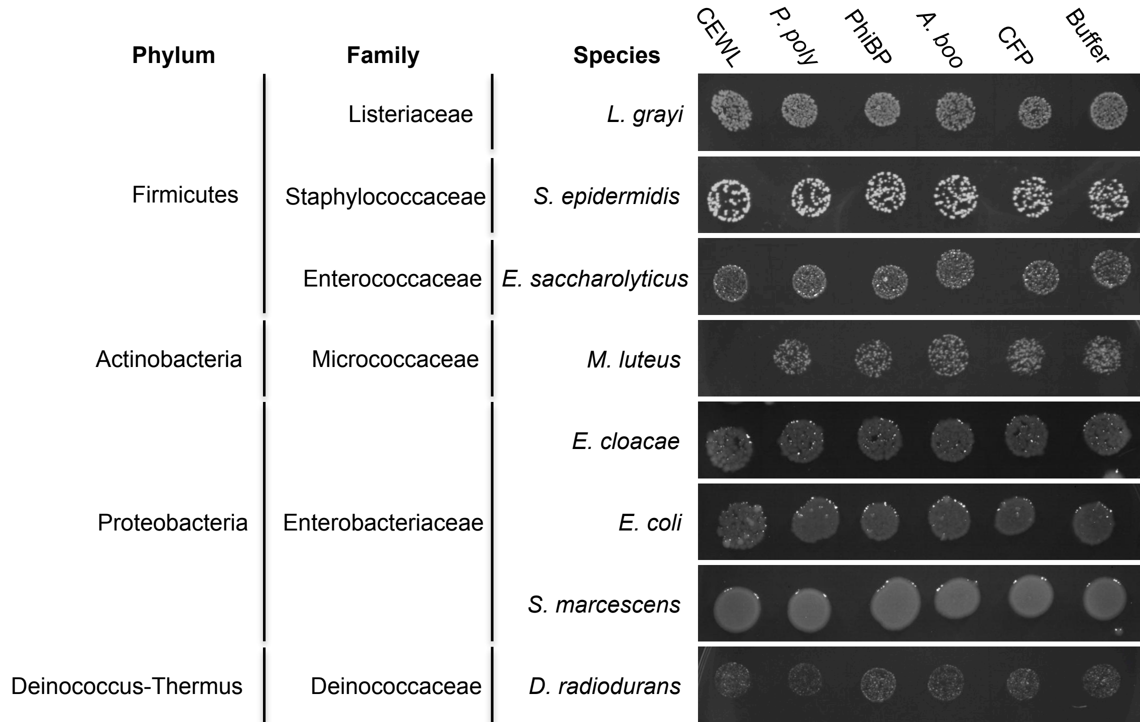


Figure III-12. Antibacterial test of *A. boonei* GH25 muramidase on additional bacteria. Bacteria of the specified strain/species incubated overnight on tryptic soy agar after a 20-minute liquid preincubation with the proteins indicated. Genera: *L* = *Listeria*, *S* = *Staphylococcus*, *E. saccharolyticus* = *Enterococcus, M* = *Micrococcus*, *E. cloacae* = *Enterobacter*, *E. coli* = *Escherichia*, *S* = *Serratia*, *D* = *Deinococcus*. Proteins: CEWL = chicken egg white lysozyme, *P. poly* = *P. polymyxa* lysozyme, PhiBP = bacteriophage PhiBP lysozyme, *A. boo* = GH25 domain of *A. boonei* lysozyme, CFP = cyan fluorescent protein. Images are representative of at least three independent experiments. Bacterial strains used are listed in Table III-3.

Table III-3. Bacterial strains used in antibacterial assays. American Type Culture Collection (ATCC) reference strain is indicated when available.

Species/strain	Source
Bacillus megaterium	Ward's Scientific
Bacillus subtilis ATCC 19659	Microbiologics, Inc.
Paenibacillus polymyxa ATCC 842	Microbiologics, Inc.
Paenibacillus polymyxa ATCC 7070	Microbiologics, Inc.
Listeria grayi ATCC 25401	Microbiologics, Inc.
Staphylococcus epidermidis ATCC 49134	Microbiologics, Inc.
Enterococcus saccharolyticus ATCC 43076	Microbiologics, Inc.
Micrococcus luteus ATCC 49732	Microbiologics, Inc.
Enterobacter cloacae	Ward's Scientific
Escherichia coli	Ward's Scientific
Serratia marcescens	Ward's Scientific
Deinococcus radiodurans ATCC 13939	Microbiologics, Inc.

Third, the *A. boonei* muramidase domain is part of a larger gene (1725 bp) composed of other domains that may broaden or constrain the range of antibacterial activity. To test the full-length gene's function in the absence of genetic tools in this system, we cloned the entire gene into an expression plasmid in *E. coli* and discovered that bacterial colonies grew poorly, with tiny, slow-growing colonies on solid media, and substantial cell death coinciding with a small amount of leaky expression in liquid culture (Figure III-13A). However, two colonies grew to normal size and upon sequencing, we determined that their expression plasmids were disrupted by insertions of 774bp (mutant 1) and 768bp (mutant 2) of a native IS1 family transposase from *E. coli* at 21bp or 266bp from the start of the lysozyme gene, respectively. These insertions also resulted in a number of premature stop codons in the lysozyme reading frame, disrupting production

of the full-length gene (Figure III-13B). Thus, *E. coli* death requires intact and full-length lysozyme, and toxicity is not due to expression construct itself. In sum, expression of the complete lysozyme resulted in *E. coli* death, while cloned genes with insertion sequences and premature stop codons abolished the lytic capacity of these proteins from within *E. coli* cells, providing evidence for an expanded host range to the antibacterial action.

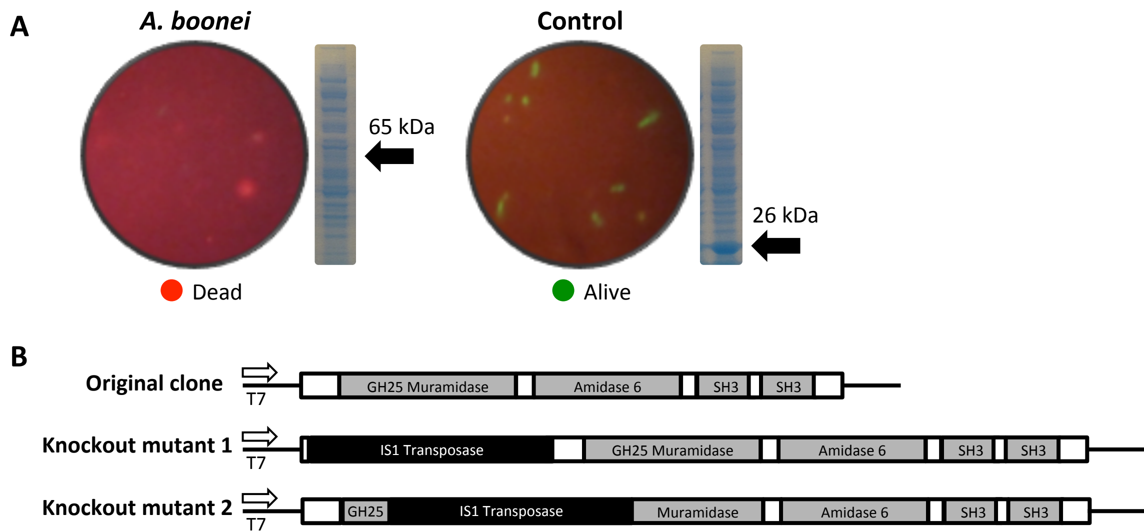


Figure III-13. *E. coli* death following full-length *A. boonei* lysozyme expression. (A) Live/dead stain of BL21 (DE3) *E. coli* transformed with expression constructs for the full-length lysozyme from *A. boonei* or WORiA, a bacteriophage infecting *Wolbachia pipientis* strain *w*Ri, as a control after overnight growth without induction. PAGE gels of crude *E. coli* lysates from *E. coli* expressing the indicated lysozyme after six hours of induction are also shown with the expected sizes of lysozymes indicated with arrows. **(B)** Structure of original full-length *A. boonei* lysozyme expression plasmid and two spontaneous knockout mutants caused by insertion of 774bp (mutant 1) and 768bp (mutant 2) of IS1 transposase sequences. Knockout mutants grew to normal colony size, while all wild type colonies had intact expression plasmids, grew poorly, and died over time in liquid culture.

Fourth, if horizontally transferred lysozymes serve as antibacterials to fend off bacterial niche competitors, two predictions follow: the lysozyme will be upregulated in response to bacterial competition and upregulation may correlate with a relative growth advantage in coculture. We thus cultured *A. boonei* cells in anaerobic marine media (Reysenbach et al., 2006) with and without cohabiting *Mesoaciditoga lauensis* (phylum Thermotogae) that was isolated from the same hydrothermal vent field as *Aciduliprofundum* in the Eastern Lau Spreading Center (Reysenbach et al., 2013). As expected, we observed a significant increase in *A. boonei* lysozyme expression at four (up to 127% increase) and twelve hours (up to 43% increase) of coculture with *M. lauensis* in comparison to negative control cultures of the singular *A. boonei* (Figure 7A). Ideally, *A. boonei* wild type and lysozyme knockouts would be employed to test relative fitness and bacterial inhibition. However, genetic manipulation of *A. boonei* is not currently feasible.

Growth experiments of *A. boonei* and *M. lauensis* were continued for 72 hours, during which there was a relative Malthusian fitness (Lenski et al., 1991) increase for *A. boonei* in coculture vs. monoculture (Figure 7B) across the exponential growth phase. This difference is marginally non-significant, perhaps due to low sample sizes ($P = 0.11$, $N = 5$, MWU two-tailed test). When the species are cultured separately for 72 hours, *M. lauensis* cell abundance is greater than that of *A. boonei* during 14 out of the 19 sampling points (Figure 7C, blue circles), indicating that bacteria outperform archaea in monoculture conditions. However, when the two species are cocultured, the cell abundances reverse and *A. boonei* outperforms *M. lauensis* for 14 out of the 19 time points (Figure 7D, red circles). This competitive frequency difference is significant (Chi-square test, $P = 0.0035$), complementing the Malthusian fitness increase. Additionally,

for each monoculture time point, there are 4.43% fewer *A. boonei* cells on average than *M. lauensis*, while in coculture there are 6.22% more *A. boonei* cells per time point (Mann Whitney U. p = 0.023). Thus, *A. boonei* outcompetes its bacterial competitor in coculture despite a higher monoculture growth rate for the bacteria.

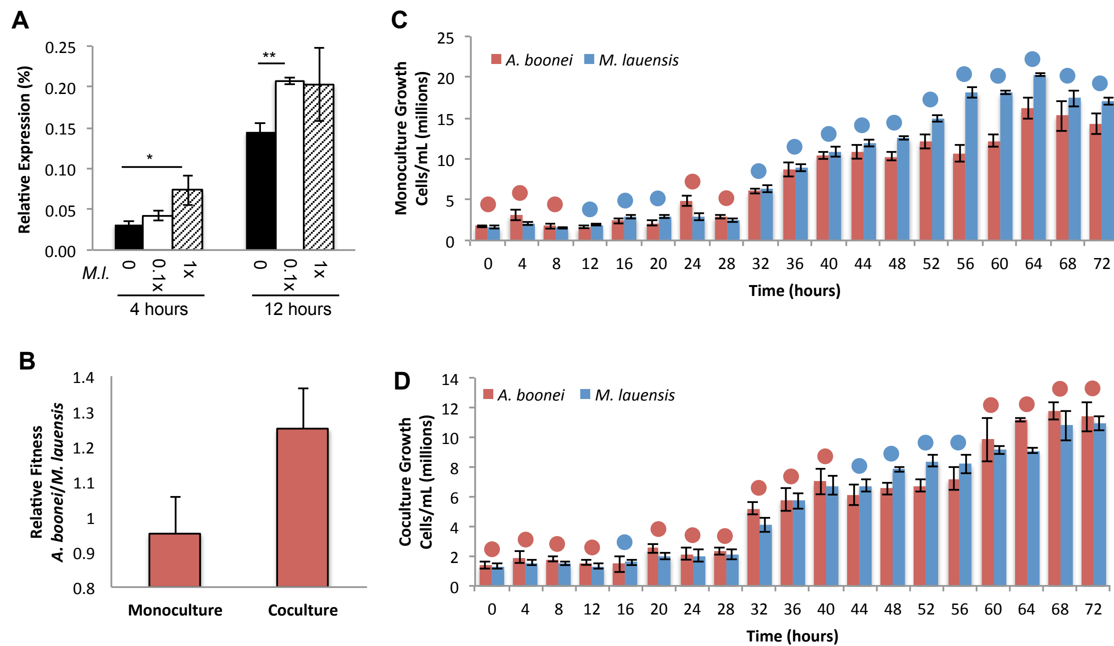


Figure III-14. Lysozyme expression and relative fitness during *A. boonei* and *M. lauensis* coculture. (A) Expression of *A. boonei* GH25 muramidase relative to the control gene elongation factor 1α, after the indicated time of coculture with *M. lauensis (M.l)* at the specified ratio relative to *A. boonei*. * P<0.05, ** P<0.01, by Mann-Whitney U pairwise comparisons. N = 6 for all samples. Primers are listed in Table S3. **(B)** Relative fitness of *A. boonei* vs. *M. lauensis* in monoculture (N=5) and coculture (N=4). **(C)** Growth of *A. boonei* (red) and *M. lauensis* (blue) monocultures over time. Significant differences in cell abundance occur at 24, 52, and 64 hours (P<0.05), and 56 and 60 hours (P<0.01) based on pairwise Wilcoxon tests. **(D)** Growth of *A. boonei* and *M. lauensis* in coculture over time. Significant differences in cell abundance occur at 48, 52, and 64 hours (P<0.05) based on pairwise Wilcoxon tests. Error bars are +/- SEM for all panels.

Conclusions

The universal tree represents the evolutionary relationships between cellular domains and establishes the modern foundation for benchmarking the magnitude of HGT across life. Indeed, HGTs have been described between each domain including archaea and bacteria (Nelson et al., 1999, van Wolferen et al., 2013), bacteria and eukaryote (Andersson, 2005, Bordenstein, 2007, Gladyshev et al., 2008, Danchin et al., 2010), and archaea and eukaryote (Schonknecht et al., 2013, Andersson et al., 2003). Despite these cases and others (Zhaxybayeva and Doolittle, 2011, Brown, 2003), HGTs are not without limits and often succumb to the selective costs of genomic rearrangements, cytotoxic effects, disruptive insertions, and functional inefficiencies upon integration (Baltrus, 2013).

It follows then that HGTs do not occur at equal rates across the universal tree, but rather experience preferential routes in which the costs of HGT are easier to overcome. The resulting pattern of HGT can be understood as a gradient of decreasing frequency from within domain > between two domains > between all domains of life (Bruto et al., 2013, Zhaxybayeva and Doolittle, 2011, Puigbo et al., 2009, Andam and Gogarten, 2011, Andam and Gogarten, 2013). In support of this pattern, the overwhelming evidence of gene transfers between bacteria is counterbalanced by the extreme lack of parallel gene transfers across all extant groups of life. As these parallel transfers are usually ancient and occur in non-model organisms (Lundin et al., 2010, Koonin et al., 2003, McClure, 2001, McDonald et al., 2012, Moran et al., 2012), they can defy clear interpretations due to their deep antiquity and lack of functional validation.

One feature that parallel HGTs have in common is that the gene's phenotype must transcend different physiological capabilities, cellular structures, and ecological niches to repeatedly increase the fitness of each recipient across the whole diversity of life. While not traditionally used in the context of parallel HGT across all cellular domains, the term niche-transcending gene appropriately captures these conditions (Wiedenbeck and Cohan, 2011). The lysozyme gene family we describe in archaea, bacteria, eukaryotes, and viruses provides one such example because the adaptive benefit of an antibacterial muramidase has repeatedly surmounted the obstacles against recurrent HGT. Indeed, horizontally transferred homologs of the GH25 muramidase exhibit differential tissue expression in *A. pisum* (Nikoh et al., 2010) and bacteriolytic activity in the fungus *Aspergillus nidulans* (AN6470.2) (Bauer et al., 2006). Thus, the horizontally transferred homologs in eukaryotes confer the same transcriptional and enzymatic activity as in the archaea.

The muramidase in a thermophilic archaea is of special note as archaea do not possess murein cell walls (Albers and Meyer, 2011), and genes encoding an antibacterial peptide have never before been identified (Cantarel et al., 2009). Members of the genus *Aciduliprofundum* are widespread thermoacidophiles in deep-sea hydrothermal vent chimney biofilms (Flores et al., 2012) in which bacteria are frequent inhabitants (Orcutt et al., 2011b, Miroshnichenko and Bonch-Osmolovskaya, 2006), including the *M. lauensis* species tested above. Archaea have been largely ignored in the context of antibiotic discovery, likely because of the conjecture that archaea do not compete with bacteria in nature. However, given that they coexist with diverse bacterial species in the environment (Oren, 2002, Kato and Watanabe, 2010, Orcutt et al., 2011b) and can

compete for similar resources, there may be significant, unexploited potential for antibiotics in this domain. Only a handful of antimicrobial peptides produced by archaea have been characterized, and those are active only against other archaea (O'Connor and Shand, 2002) despite the fact that archaea are known to inhibit bacteria in diverse environments (Atanasova et al., 2013, Shand, 2008). It is also possible that since *Aciduliprofundum* strains metabolize peptides, the lysozyme enables a nutritive strategy in which lysed bacteria provide nutrients for the archaeon to scavenge.

Based on this work, we suspect that systematic surveys of archaea gene products will likely uncover a broad range of antibacterial activities, and may eventually offer novel peptide or small molecule therapeutics. Such antibacterial products may have naturally evolved thermostability that would increase their attractiveness as therapeutics. GH25 muramidases have been demonstrated as effective antibacterials against biofilms of *Streptococcus pneumoniae* (Domenech et al., 2011) and related enzymes have proven efficacious in mouse models of bacterial mucosal colonization (Fenton et al., 2010), sepsis (Loeffler et al., 2003), and endocarditis (Entenza et al., 2005).

In summary, we infer that the evolutionary path to this parallel HGT was paved by the universal drive for nonbacterial taxa to compete in a bacterial world. We predict that similar to the cascade of antibiotic gene transfer discoveries that followed their initial reporting, parallel transfers of genes to all cellular domains and viruses might regularly have antimicrobial functions.

Materials and methods

Unless otherwise stated, reagents were obtained from Fisher Scientific (Waltham, WI).

PCR and sequencing

PCR was performed using GoTaq DNA Polymerase (Promega, Madison, WI) with primers listed in Figure 1 – figure supplement 2. PCR products were electrophoresed using 1% agarose gels in sodium boric acid buffer. Following electrophoresis, gels were dyed with GelRed (Phenix Research, Candler, NC) and imaged on an Alpha Innotech GelRed Imager (Alpha Innotech, San Leandro, CA). Amplified bands were excised from the gels and purified with an SV Wizard Gel Cleanup kit (Promega). Following purification, DNA concentration was measured using the Qubit DNA high sensitivity kit (Life Technologies, Grand Island, NY) and sequencing reactions were performed by Genewiz (South Plainfield, NJ).

Bioinformatics

The lysozyme protein from *Wolbachia* prophage WORiA (ZP_00372884) was used as a query in a blastp search of the NCBI nonredundant protein database using Geneious Pro v5.5.6. All hits with E-values below 10^{-12} were collected and duplicate entries were removed. Sequences from field and laboratory samples were added to this collection and aligned with MUSCLE (Edgar, 2004), insertions and deletions were removed, and the 8 most highly conserved residues from the MUSCLE alignment were mapped to a structure prediction of *A. boonei* lysozyme using PyMOL. Structure

prediction was performed using the homology-based modeling tool Phyre2 (Kelley and Sternberg, 2009). For phylogenetic analyses, ProtTest (Abascal et al., 2005) was used to determine the best model of protein evolution based on the corrected Akaike information criterion (AICc). MrBayes (Ronquist et al., 2012) and PhyML (Guindon et al., 2010) were used to build a phylogenetic tree with Bayesian and maximum likelihood methods, respectively. For the global lysozyme phylogeny, the best model chosen by ProtTest (LG + I + G) was used to generate the maximum likelihood tree, while the 3rd best model (WAG + I + G; ΔAICc: 74.82) was used to generate the Bayesian tree due to a lack of LG model availability in MrBayes. *S. sanguinolenta* and *S. stauntoniana* lysozymes were excluded from this analysis because frameshift mutations suggest the genes may be evolving in the absence of selection, while Aphidinae lysozymes were not included because of shorter sequences of the GH25 muramidase domain obtained through the use of degenerate primers that would have limited resolution of the tree.

In an iterative approach, each candidate example of HGT was used as a blastp query against the nr database and the top 15 (*A. boonei*, *A. pisum*, *S. moellendorffii*) or top 75 (*A. oryzae*) E-value hits were subjected to the same phylogenetic analysis as above. Evolutionary models used were *A. boonei*: WAG + I + G on a 156aa indel-free alignment, *A. pisum*: CpREV + I + G on a 190aa indel-free alignment, *S. moellendorffii*: WAG + I + G (ΔAICc: 44.28) on a 200aa indel-free alignment, *A. oryzae*: WAG + G (Bayesian, ΔAICc: 4.49) or LG + G (maximum likelihood) on a 186aa indel-free alignment. The fungal lysozyme was also phylogenetically analyzed on the DNA level using the top 25 E-value blastn hits to exon 2 of the *A. oryzae* lysozyme gene. jModelTest

2 (Darriba et al., 2012) was used to determine the best model of nucleic acid evolution (GTR + I + G, ΔAICc: 7.33) of a 282bp indel-free alignment. The Archaea HGT clade was also analyzed phylogenetically with a Bayesian tree of selected taxa using lysozyme protein sequences (WAG + I + G, 185aa) and compared to 16S rRNA (GTR + G, 1,156bp) for the same strains obtained from SILVA (Quast et al., 2013).

Statistical support for the HGT hypothesis was assessed with the Shimodaira-Hasegawa test (SH-test) (Shimodaira and Hasegawa, 1999) as implemented in RAxML v.8.0.20 (Stamatakis, 2014). An unresolved binary constraint tree was generated in MacClade v4.08, in which bacterial sequences are monophyletic, as are nonbacterial sequences, with all other topology unconstrained. This constraint tree was used to generate a maximum likelihood best tree with RAxML, using the same evolutionary models as above. The SH-test was then run comparing the maximum likelihood constrained tree to the unconstrained consensus Bayesian tree or to 100 bootstrap trees from the maximum likelihood analysis from PhyML

Lysozyme cloning and purification

A. boonei GH25 muramidase domain (ZP_04874596), *P. polymyxa* lysozyme (YP_003869492), and PhiBP lysozyme (CBA18122) were cloned and expressed with a 6x C-terminal histidine tag using an Expresso T7 Cloning and Expression System (Lucigen, Middleton, WI) according to the manufacturers instructions. We also cloned the *S. moellendorffii* and *A. oryzae* GH25 muramidases, however recombinant 6x histidine-tagged proteins were insoluble when expressed in either *E. coli* or sf9 insect

cells and attempts to solubilize them were unsuccessful. Sequence-confirmed expression plasmids and a control plasmid expressing cyan fluorescent protein (CFP) were transformed into HI-Control BL21 (DE3) *E. coli* cells. Cultures at an OD600 of ~0.5 were induced with 1 mM IPTG for 6 hours, centrifuged, and frozen at -80 ˚C until purification. Frozen pellets were resuspended in lysis buffer containing 10 mM Tris-HCl, pH 7.5, 300 mM NaCl, 0.5% Triton x-100, 0.3% sodium dodecyl sulfate, and 1 mM phenylmethylsulfonylfluoride and sonicated 5 times for 30 seconds with at least 1 minute on ice between sonications. Samples were centrifuged and recombinant proteins were purified from supernatant using HisPur Ni-NTA chromatography cartridges (Thermo Scientific, Waltham, MA) according to manufacturer's instructions. Glycerol at a final concentration of 40% was added to enzymes in elution buffer for storage at -20 °C for a maximum of three weeks before use in antibacterial assays. Purifications were analyzed with denaturing polyacrylamide gel electrophoresis and stained with GelCode Blue (Thermo Scientific).

Full-length *A. boonei* lysozyme and WORiA lysozyme were cloned into a pET-20b vector (EMD Millipore, Darmstadt, Germany) with a C-terminal 6x histidine tag and sequence-confirmed plasmids were transformed into BL21 (DE3) *E. coli* (EMD Millipore). Three colonies from each transformation were inoculated into LB media and grown to an OD600 of ~0.5, induced for 4 hours with 1 mM IPTG and harvested for analysis on PAGE gels. Overnight cultures without induction were examined for bacterial death with a BacLight Live/Dead Stain (Life Technologies).

Antibacterial assays

Purified *A. boonei* GH25 muramidase, *P. polymyxa* lysozyme, PhiBP lysozyme, CFP, and commercially purchased CEWL (Sigma-Aldrich, St. Louis, MO) were diluted to 100 μg/mL in buffer EG (60% nickel column elution buffer, 40% glycerol) and filter sterilized. Bacteria to be tested were grown overnight in tryptic soy broth, split 1:10, and incubated to exponential growth before being diluted into each enzyme solution. Samples were incubated with shaking for 20 minutes at 37 °C and then 5 μL was spotted onto tryptic soy agar and incubated overnight at 37 °C. To evaluate whether antibacterial activity is dose-dependent, *B. subtilis* was incubated with *A. boonei* GH25 muramidase at 100 μg/mL, 75 μg/mL, 50 μg/mL, 25 μg/mL and 0 μg/mL and 100 μl was spread on tryptic soy agar plates. Replicates of 10 were performed for each concentration, plates were incubated overnight at 37 °C, and colonies were counted the following morning. Bacterial strains used in these experiments are listed in Figure 5 – figure supplement 3.

A. boonei cultures

A. boonei and *M. lauensis* cultures were performed as previously described (Reysenbach et al., 2006) with the following modifications: yeast extract was added at 2.0 g/L, pH was adjusted to 4.8, and cultures were incubated at 65 °C. For gene expression studies, 8.2 x 10^5 cells were inoculated into 5 mL cultures in 6 replicates each of monocultures and cocultures at 0.1:1, 1:1, and 1:0.1 ratios and 500 μL samples were collected after 4 and 12 hours of co-incubation and frozen for expression analysis. RNA was isolated from frozen samples using an RNeasy Mini Kit (Qiagen) and QIAshredder

(Qiagen), DNA contamination was removed with a Turbo DNAfree Kit (Life Technologies), and reverse transcription was performed using a Superscript III 1st Strand Synthesis System (Life Technologies) along with no-reverse transcriptase controls. Quantitative PCR was performed with GoTaq qPCR Master Mix (Promega) using a CFX96 Real-Time System (Bio-Rad, Hercules, CA). Primers are listed in Table S3. For competition studies, 5 replicates of 5 mL cultures were inoculated as monocultures or 1:1 cocultures and 175 μL was collected every 4 hours for counting of relative species abundance with a hemocytometer. Relative fitness was calculated based on Malthusian parameters over the period of exponential growth as previously described (Lenski et al., 1991).

CHAPTER IV. RECENT GENOME REDUCTION OF *WOLBACHIA* IN *DROSOPHILA RECENS* TARGETS PHAGE WO AND NARROWS CANDIDATES FOR REPRODUCTIVE PARASITISM[‡]

Abstract

Wolbachia are maternally transmitted endosymbionts that often alter their arthropod hosts' biology to favor the success of infected females, and may also serve as a speciation microbe driving reproductive isolation. Two of these host manipulations include killing males outright or causing markedly reduced offspring survival when infected males mate with uninfected females, a phenomenon known as cytoplasmic incompatibility. Little is known about the mechanisms behind these phenotypes, but interestingly either effect can be caused by the same *Wolbachia* strain when infecting different hosts. For instance, *w*Rec causes cytoplasmic incompatibility in its native host *Drosophila recens* and male killing in *D. subquinaria*. The discovery of prophage WO elements in most arthropod *Wolbachia* has generated the hypothesis that WO may encode genes involved in these reproductive manipulations. However, PCR screens for the WO minor capsid gene indicated that *w*Rec lacks phage WO. Thus, *w*Rec seemed to provide an example where phage WO is not needed for *Wolbachia*-induced reproductive manipulation. To enable investigation of the mechanism of phenotype switching in different host backgrounds, and to examine the unexpected absence of phage WO, we sequenced the genome of *w*Rec. Analyses reveal that *w*Rec diverged from *w*Mel approximately 350,000 years ago,

---

[‡] This chapter is published in *PeerJ*, 2014, 2:e529. Minhee Jo, Sarah R. Bordenstein, John Jaenike, and Seth R. Bordenstein contributed to the authorship of this chapter.

mainly by genome reduction in the phage regions. While it lost the minor capsid gene used in standard PCR screens for phage WO, it retained two phage WO regions encompassing 33 genes, several of which have previously been associated with reproductive parasitism. Thus, WO gene involvement in reproductive manipulation cannot be excluded and reliance on single gene PCR should not be used to rule out the presence of phage WO in *Wolbachia*. Additionally, the genome sequence for *w*Rec will enable transcriptomic and proteomic studies that may help elucidate the *Wolbachia* mechanisms of altered reproductive manipulations associated with host switching, perhaps among the 33 remaining phage genes.

Introduction

Wolbachia are widespread obligate intracellular α-proteobacteria that infect around 40% of arthropod species (Zug and Hammerstein, 2012) and 47% of filarial nematodes (Ferri et al., 2011). These infection frequencies, if extrapolated to the diversity and abundance of their hosts, make *Wolbachia* perhaps the most widespread endosymbiont in animals. To maximize its propagation in arthropods, the maternally inherited *Wolbachia* has evolved an assortment of mechanisms to distort its host's reproductive system in a manner that enhances the relative production of infected females. These mechanisms include feminization, parthenogenesis, male killing, and cytoplasmic incompatibility (CI), the most common phenotype and one that results in embryonic lethality when matings occur between infected males and uninfected females (Werren et al., 2008). Females harboring the same *Wolbachia* strain, meanwhile, can successfully mate and

produce infected offspring with either infected or uninfected males, giving these females a selective advantage in populations of mixed infection status.

Interestingly, some *Wolbachia* strains are multipotent and can induce more than one type of reproductive manipulation depending on the arthropod host it infects (Fujii et al., 2001, Jaenike, 2007). In one striking example, the *Wolbachia* strain *w*Rec causes CI in its native host, *Drosophila recens*, but when introgressed into a sibling species, *D. subquinaria,* it causes male killing (Jaenike, 2007). Moreover in a natural hybrid zone between these same two species, unidirectional CI plays a major role in reducing interbreeding and thus contributes to reproductive isolation between these species (Jaenike et al., 2006, Shoemaker et al., 1999). Even though the link between *Wolbachia* and CI has been known for over 40 years (Yen and Barr, 1971), the mechanisms by which *Wolbachia* accomplishes its reproductive manipulations remain unknown. Despite the physical isolation resulting from its intracellular lifestyle, *Wolbachia* in arthropods are replete with mobile DNA (Wu et al., 2004) including a temperate bacteriophage named WO (Kent and Bordenstein, 2010, Metcalf and Bordenstein, 2012, Masui et al., 2000). It has been speculated that WO may be involved in *Wolbachia* reproductive manipulations due to the prevalence of ankyrin repeat genes in its genome (Wu et al., 2004), the pervasiveness of phage-encoded bacterial virulence factors (Boyd, 2012), and the frequent occurrence of phage WO in arthropod *Wolbachia* strains (Gavotte et al., 2007). However, evidence not supportive of this hypothesis includes the observations that CI is inconsistently associated with the presence of phage WO genes (Sanogo et al., 2005, Saridaki et al., 2011) and that the penetrance of CI in *Nasonia* wasps is negatively correlated with densities of phage WO virions, as predicted by the phage density model

(Bordenstein et al., 2006, Bordenstein and Bordenstein, 2011). Interestingly, PCR screening for the WO minor capsid gene specified WO's absence in *w*Rec (Bordenstein and Wernegreen, 2004), even though its closest relatives have large amounts of prophage DNA (Wu et al., 2004, Klasson et al., 2009b). Thus, the absence of phage WO in *w*Rec would be a critical example of a *Wolbachia* strain causing multiple reproductive phenotypes but lacking WO.

To investigate the apparent lack of prophage WO genes and alternative genetic mechanisms behind *w*Rec's diverse phenotypic influences, we sequenced the *w*Rec genome using next-generation sequencing technology with partial finishing via Sanger sequencing. We determined that although *w*Rec lacks the WO minor capsid gene typically used in diagnostic screens, it does contain a number of prophage WO genes. Thus, the possibility that WO influences *Wolbachia* reproductive manipulations cannot be eliminated, and those WO genes present in *w*Rec offer a streamlined candidate list of the WO genes that could cause reproductive parasitism. Additionally, the availability of genomic information for a *Wolbachia* strain that is known to switch reproductive phenotypes will enable genomic, transcriptomic, and proteomic approaches to investigate the mechanisms behind these phenotypes.

Materials & Methods

The *w*Rec genome sequencing reads and annotated contigs can be accessed from NCBI Bioproject PRJNA254527.

Wolbachia strain relatedness

Multi-locus sequence typing (MLST) genes were concatenated and a Bayesian phylogeny was inferred as previously described (Baldo et al., 2006). Briefly, selected fragments of MLST genes (coxA, gatB, fbpA, ftsZ, and hcpA) from *Wolbachia* strains with complete or nearly complete genome sequences were obtained from GenBank or the sequencing group's online repositories, concatenated for a total length of 2,079bp, and aligned with MUSCLE (Edgar, 2004). jModelTest 2 (Darriba et al., 2012) was used to determine the best model of evolution for the set of MLST haplotypes (GTR + I + G), and a Bayesian phylogeny was inferred using Mr. Bayes (Ronquist et al., 2012) with a chain length of 1,100,000, burn-in of 100,000, and subsampling frequency of 200.

Genome sequencing and assembly

DNA was extracted from a pool of 10 female *Wolbachia* infected Pittsford strain *D. recens* flies using a Puregene DNA purification kit (Qiagen, Venlo, Limburg). Shotgun sequencing of the *w*Rec genome was conducted using an Illumina Hi-Seq (Vanderbilt Sequencing Core, Nashville, TN) with 100 bp paired end sequencing. Reads were filtered using five available *Wolbachia* genomes, *w*Bm (Foster et al., 2005), *w*Mel (Wu et al., 2004), *w*Ri (Klasson et al., 2009b), *w*Oo (Darby et al., 2012), and *w*Pip (Klasson et al., 2008), by mapping reads to these genomes with length and similarity fractions of 0.5 and keeping all mapped reads, using CLC Genomics Workbench version 6.0.4 (CLC Inc, Aarhus, Denmark). A *de novo* assembly with a length fraction of 0.5 and similarity fraction of 0.8 was then performed on filtered reads. Sequencing of whole *w*Rec-infected *D. recens* females produced over 24 million reads, of which nearly 4%

matched one or more previously sequenced *Wolbachia* genomes. *De novo* assembly of *Wolbachia*-filtered reads yielded 159 contigs. A *de novo* assembly of unfiltered reads was also performed and any contigs with a portion of its sequence matching contigs obtained from the filtered assembly were added to scaffolds in a search for novel genes. Separately, reads were mapped to the *w*Mel genome with length and similarity fractions of 0.5 producing a rough consensus sequence to guide assembly of the de novo contigs into scaffolds, which were further refined with Sanger sequencing of PCR amplifications using primers designed to bind either end of putatively adjacent contigs to yield a final draft genome consisting of 43 scaffolds.

Annotation and comparative genomics

The *w*Rec genome was annotated using MicroScope (Vallenet et al., 2009), supplemented with manual curation based on homology with *w*Mel. A comparison of gene-gene identity between *w*Mel and *w*Rec was performed with a reciprocal best BLAST as previously described (Moreno-Hagelsieb and Latimer, 2008). For whole-genome alignments and analyses, *w*Rec scaffolds were concatenated in the order in which the majority of their genes appear in *w*Mel. Whole-genome alignment was performed with Mauve (Darling et al., 2010) and a circular genome plot was created with DNAPlotter (Carver et al., 2009). Manual annotations, BLAST searches, and sequence manipulation were performed with either CLC Genomics Workbench or Geneious V5.5.6 (Biomatters Ltd., Auckland, New Zealand). Ka/Ks rates and ratios were calculated using either single gene or concatenated whole genome CDS alignments with any alignments shorter than 30 amino acids removed (Buschiazzo et al., 2012), using the program DnaSP

(Librado and Rozas, 2009). Genomic synteny was assessed with the Cloud Virtual Resource (CloVR) comparative pipeline (Angiuoli et al., 2011) and Sybil synteny gradient viewer (Riley et al., 2012) using the Data Intensive Academic Grid (DIAG) at the University of Maryland. The number of phage and phage-associated genes in *Wolbachia* genomes was determined based on current GenBank annotations and includes genes in the phage-packaged eukaryotic association module (Sarah Bordenstein, unpublished data).

Results

Taxonomy of wRec

Phylogenetic analysis based on the concatenated multilocus sequence typing (MLST) genes (Baldo et al., 2006) confirms several previous reports that the supergroup A strain *w*Rec is closely related to *w*Mel (Baldo et al., 2006, Werren et al., 1995, Ioannidis et al., 2007, Gueguen et al., 2012), a widespread strain infecting *D. melanogaster* (Figure IV-1). In addition, phylogenetic analyses of each individual MLST gene support the same relationship of *w*Mel as the closest sequenced relative to *w*Rec (data not shown). To date, all sequenced *Wolbachia* genomes in supergroups A and B, including *w*Mel (Wu et al., 2004), have contained significant amounts of phage WO DNA. Thus the potential absence of WO in *w*Rec was unexpected and precipitated the genomic analysis described below.

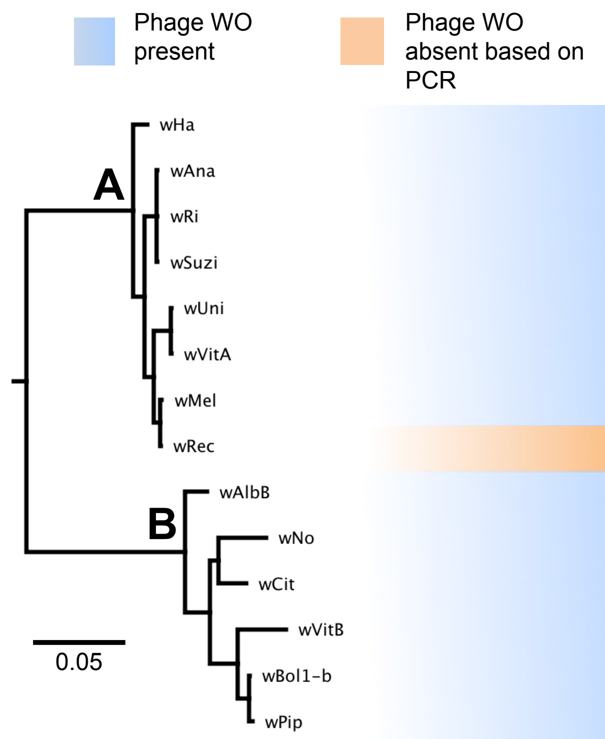


Figure IV-1. WO phage is present in all sequenced supergroup (A) and (B) *Wolbachia* strains. A Bayesian phylogeny based on the concatenated *Wolbachia* multi-locus sequence typing genes is shown, consisting of selected strains with partial or full genome sequences and wRec. All branches had posterior probabilities of 99% or greater. While all previously sequenced *Wolbachia* strains in group (A) and (B) possess phage WO elements, wRec (indicated with arrow), was formerly thought to be phage-free.

Genome features of wRec with targeted reduction of prophage WO

Full sequencing statistics and an overview of *w*Rec genome features are listed in Table IV-1. *w*Rec scaffolds (N=43) consisted of a total sequence length of 1,126,653 bp containing 1271 protein coding sequences. 99.7% of all nucleotides in coding sequences shared between *w*Rec and *w*Mel were identical, indicating little divergence between these two closely related genomes despite occupying hosts that diverged >50 million years ago (Ross et al., 2003). Based on a previously established rate of synonymous substitution in *Wolbachia* of 0.9% per million years (Raychoudhury et al., 2009), the genome-wide

percentage of synonymous substitution (0.314%) between *w*Rec and *w*Mel puts their divergence at approximately 350,000 years ago. There were 2,009 single nucleotide polymorphisms (SNPs) between shared coding genes in *w*Mel and *w*Rec, and 599 (29.8%) of these SNPs were synonymous with an average K_a/K_s ratio for each gene of 0.691. The vast majority of genes are highly conserved between *w*Rec and *w*Mel. More than 95% of orthologous gene pairs were 99% identical or greater and only ten gene pairs were less than 98% identical (Table IV-2). Most of these divergent genes code for hypothetical proteins and ankyrin repeat domain proteins. The *wsp* surface antigen, a known hypervariable sequence in *Wolbachia* (Zhou et al., 1998), was also among the less conserved loci. All ten divergent genes contained insertions or deletions compared to *w*Mel in addition to one or more SNPs. Interestingly, four of these divergent genes, two coding for hypothetical proteins, an Ovarian Tumor (OTU)-like cysteine protease, and *wsp*, had K_a/K_s ratios greater than one (Table IV-2), suggesting that they are evolving under positive selection, and the proteins they encode may be relevant to strain-specific host interactions. When these four genes were aligned to their homologs in *w*VitA, the closest relative of *w*Mel and *w*Rec, a roughly equal number of mutations in the OTU protease and *wsp* genes in each strain matched the sequence in *w*VitA. However, for the two hypothetical proteins WREC_0649 (WD_0722) and WREC_1268 (WD_1278), the *w*Mel alleles matched *w*VitA in a majority of cases (18 out of 25 nucleotides and 49 out of 56 nucleotides, respectively), suggesting that the *w*Mel variants were ancestral and that these *w*Rec alleles experienced lineage-specific positive selection during *D. recens* infection.

Table IV-1. *w*Rec sequencing and genome statistics.

Reads	24,633,972
*w*Rec reads	955,730 (3.9%)
Contigs	159
Scaffolds	43
Average coverage	76.5
Genome size	>1,126,653 bp
GC content	35.2%
CDS on scaffolds	1271
Average CDS length	764 bp
Average intergenic length	130 bp
Transfer RNA's	34
Ribosomal RNA's	3 (23S, 16S, 5S)
Prophage regions	2

Table IV-2. *w*Mel genes with less than 98% nucleotide identity to their orthologs in *w*Rec. Genes with a Ka/Ks ratio greater than one are highlighted.

*w*Mel Locus	*w*Rec Locus	Function	Pairwise Identity (%)	*w*Rec length	*w*Mel length	# SNPs	# Non-synonymous SNPs	Ka/Ks Ratio	Other changes
WD_0294	WREC_0283	Ankyrin Repeat Domain Protein	89.4	1815	1626	4	4	-	189 bp insertion
WD_0443	WREC_0442	OTU-like Cysteine Protease	97.1	927	906	7	6	1.59	21 bp insertion
WD_0550	WREC_0541	Ankyrin Repeat Domain Protein *TM Domains	87.4	789	990	2	2	-	99 bp deletion, C-terminal frameshift, alternate start/stop sites
WD_0722	WREC_0649	Hypothetical Protein *TM Domains	92.0	462	450	25	21	4.25	9 bp insertion, 3 bp insertion
WD_0996	WREC_0956	Transposase	89.1	744	801	1	0	0	alternate start site, transposase insertion
WD_1007	WREC_0973	Hypothetical Protein	95.1	366	351	3	2	0.42	15 bp insertion
WD_1039	WREC_1007	Collagen Triple Helix Repeat Protein	97.5	405	1425	1	1	-	9 bp insertion, scaffold break
WD_1063	WREC_1036	Wsp Surface Antigen	97.9	708	714	9	8	2.55	6 bp deletion
WD_1278	WREC_1268	Hypothetical Protein *TM Domain	92.1	2604	2766	56	51	6.07	162 bp deletion
WD_1298	WREC_1289	RpoD	97.2	1974	1929	10	6	0.39	18 bp insertion, 27 bp insertion

Interestingly, there were only two *w*Rec genes without nucleotide homology to genes in *w*Mel, even when contigs from a de novo assembly of raw host/*Wolbachia* reads were mapped to scaffolds in a search for additional genes. These two genes, WREC_0318 and WREC_0319, are hypothetical proteins with >95% nucleotide identity to sequences in two other *Wolbachia* strains, *w*Ri and *w*Ha. Meanwhile, *w*Rec lacked any homologs of 43 *w*Mel genes (Table IV-3), all but one of which are phage-related (phage genes discussed below). The single non-phage gene without homology in *w*Rec is WD_0032, which codes for a hypothetical protein with 96% similarity to the C-terminus of an ankyrin repeat-containing siRNA binding protein in *w*Ri. As is the case for many *Wolbachia* genomes, repetitive elements such as transposases and reverse transcriptases are abundant in *w*Rec and have hampered closing of the genome. 77 such repetitive genes were found in *w*Rec, and often appeared at the boundaries of scaffolds (Figure IV-2). Although genomic rearrangement between the genomes cannot be completely assessed because the *w*Rec genome is not closed, genes in *w*Rec scaffolds were universally syntenic compared to *w*Mel (Figure IV-3), with the exception of a 5kb region containing WD_0042-WD_0051 (WREC_0853-WREC_0863), consisting of repetitive transposases, reverse transcriptases, hypothetical proteins, and pseudogenes. This region would have been located on the first *w*Rec scaffold if syntenic, but instead is on scaffold 31 (Figure IV-3).

Table IV-3. *w*Mel genes with no homologs in *w*Rec. *w*Mel genes without BLASTn hits to assembled *w*Rec scaffolds (E-value < 10^{-10}) and also absent when raw *w*Rec/host sequencing reads were mapped to these genes as reference templates.

Locus Tag	*Gene*
Non-phage	
WD_0032	Hypothetical Protein
WO-A	
WD_0254	Transcriptional Regulator, Putative
WD_0256	Hypothetical Protein
WD_0257	DUF2466, Truncation
WD_0259	Conserved Hypothetical Protein
WD_0261	Conserved Hypothetical Protein, Interruption-N
WD_0262	Conserved Hypothetical Protein, Interruption-C
WD_0263	Prophage LambdaW1, DNA Methylase
WD_0264	Conserved Hypothetical Protein
WD_0265	Prophage LambdaW1, Terminase Large Subunit, Putative
WD_0266	gpW
WD_0267	Hypothetical Protein
WD_0273	Conserved Hypothetical Protein
Octomom	
WD_0512	Hypothetical Protein
WD_0513	Hypothetical Protein
WD_0514	Ankyrin Repeat Domain Protein
WO-B	
WD_0564	Hypothetical Protein
WD_0565	Patatin Family Protein
WD_0566	Ankyrin Repeat Domain Protein
WD_0567	Prophage P2W3, Tail Protein D, Putative
WD_0568	Prophage P2W3, Tail Protein X, Putative
WD_0569	Prophage P2W3, Tail Protein U, Putative
WD_0570	Prophage P2W3, Tail Tape Measure Protein, Truncated
WD_0571	Prophage P2W3, Tail Tape Measure Protein, Truncated
WD_0572	Tail Chaperone G/GT
WD_0573	RelE Pseudogene
WD_0574	Prophage P2W3, Contractile Tail Tube Protein
WD_0576	Hypothetical Protein
WD_0577	Hypothetical Protein
WD_0578	Hypothetical Protein
WD_0579	Hypothetical Protein
WD_0580	Hypothetical Protein
WD_0581	Hypothetical Protein
WD_0582	Regulatory Protein RepA, Putative
WD_0583	Conserved Hypothetical Protein, AAA_25
WD_0584	Hypothetical Protein
WD_0585	Conserved Hypothetical Protein
WD_0586	Hypothetical Protein
WD_0589	Conserved Hypothetical Protein, AAA_25
WD_0590	Conserved Hypothetical Protein, AAA_24
WD_0591	Sigma 70
WD_0594	Prophage LambdaW4, DNA Methylase
WD_0595	Conserved Hypothetical Protein

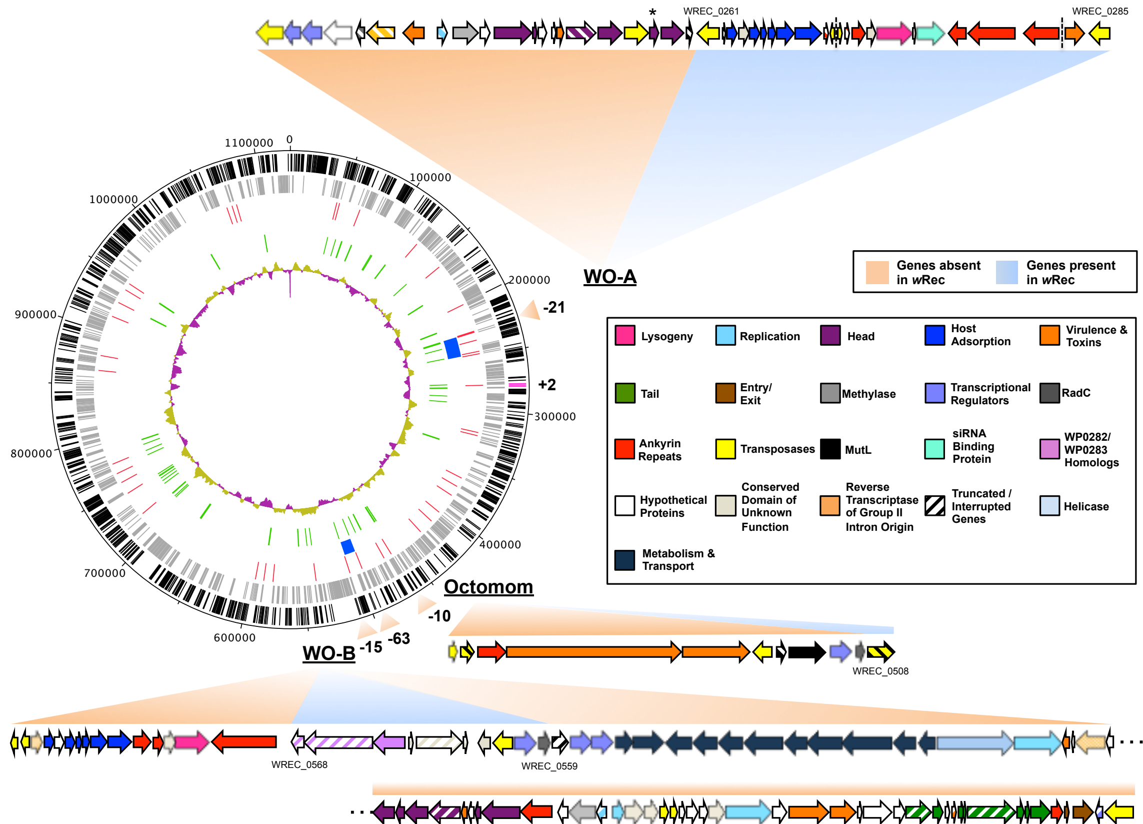


Figure IV-2. *w*Rec genome comparison to *w*Mel. *w*Rec scaffolds were concatenated in the order in which their genes appear in *w*Mel to produce the circular genome above. Major regions of loss or gain compared to *w*Mel are indicated outside the circle along with the number of genes involved. *w*Rec genome features are indicated within the circle plot as follows (from outside-in): 1 (black): CDS in forward direction, and (magenta) genes not found in *w*Mel; 2 (grey): CDS in reverse direction; 3 (red): scaffold break points; 4 (blue): WO regions; 5 (green): transposases and reverse transcriptases; 6 (purple/gold): GC content variation from average. WO prophage and related regions are shown and genes are categorized by color according to their likely functions and presence/absence in *w*Rec. Locus tags for selected genes are indicated and dashed lines indicates breaks between scaffolds containing WO-A. The minor capsid gene of WO-A, which was used for prior PCR screens, is indicated with an asterisk.

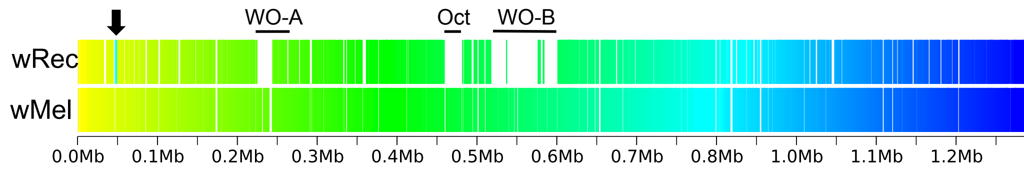


Figure IV-3. Within-scaffold *w*Rec synteny compared to *w*Mel. *w*Rec scaffolds were concatenated in the order in which they appear in *w*Mel and within-scaffold synteny was analyzed. Genes are graphed as tick marks colored on a gradient from yellow to blue from left to right with *w*Mel as the reference genome and each *w*Rec gene colored according to the location of its homolog in the *w*Mel genome. White spaces in *w*Rec alone indicate the absence of homologous genes or genes with multiple paralogs whose synteny cannot be established, while white spaces shared by both genomes indicate intergenic regions. A 5kb region of rearrangement consisting of repetitive elements and hypothetical proteins is noted with an arrow, and phage-related regions are marked.

Prophage WO relics in the genome

Whole-genome alignment of *w*Rec and *w*Mel revealed three major regions of genome reduction, with *w*Rec lacking a large portion of both phage WO regions present in *w*Mel as well as the entirety of the "Octomom" region (Chrostek et al., 2013) (Figure IV-2), with only a bordering reverse transcriptase, WREC_0508 (*w*Mel homolog WD_0506) present. Interestingly, although the minor capsid gene used in prior PCR surveys is absent, *w*Rec does contain two major phage-related regions (Figure IV-2). The first is a 19.2kb region (WREC_0261-WREC_0285) across three scaffolds that is homologous to 21 contiguous genes of *w*Mel WO-A (WD_0276-WD_0296). This region in *w*Rec is syntenic and 99.4% identical to its homologous region in *w*Mel, with two exceptions. The *w*Rec homolog (WREC_0270/WREC_0274) of WD_0285, an ankyrin repeat protein, is fragmented by the insertion of two reverse transcriptases and a gap in the scaffolds, and there is an 189bp insertion in WREC_0283 (WD_0294), another ankyrin repeat protein. If these two regions are included in the calculation of similarity,

then the *w*Rec WO-A phage region is 90.9% identical to the same region in *w*Mel. The second *w*Rec phage region contains 11.3kb and 7 genes (WREC_0559-WREC_0568) that are syntenic and homologous to part of WO-B in *w*Mel (WD_0625-WD_0632), with 99.5% pairwise identity. Two of these genes are interrupted by premature stop codons and the remaining fragments are annotated as smaller, separate genes. These genes include the orthologs of WD_0630, a hypothetical protein, which is split into three genes in *w*Rec (WREC_0563-WREC_0565), and the ortholog of WD_0632, which is split into the 3,096bp gene WREC_0567 and 468bp gene WREC_0568. While the head region of WO appears to be absent in *w*Rec, the host adsorption module which is putatively involved in binding to the host surface during phage infection is largely intact, as are a number of ankyrin repeat genes, a transcriptional regulator, and the homologs of WP_02082/WP_0283, two genes in *w*Pip recently proposed as candidate mediators of CI (Beckmann and Fallon, 2013). In summary, the markedly reduced number of phage genes in *w*Rec (N=33) is the signature feature of the genome compared to its closest relatives, which possess anywhere from 134 (*w*Ha) to 225 (*w*Ri) phage or phage-associated genes (Figure IV-4).

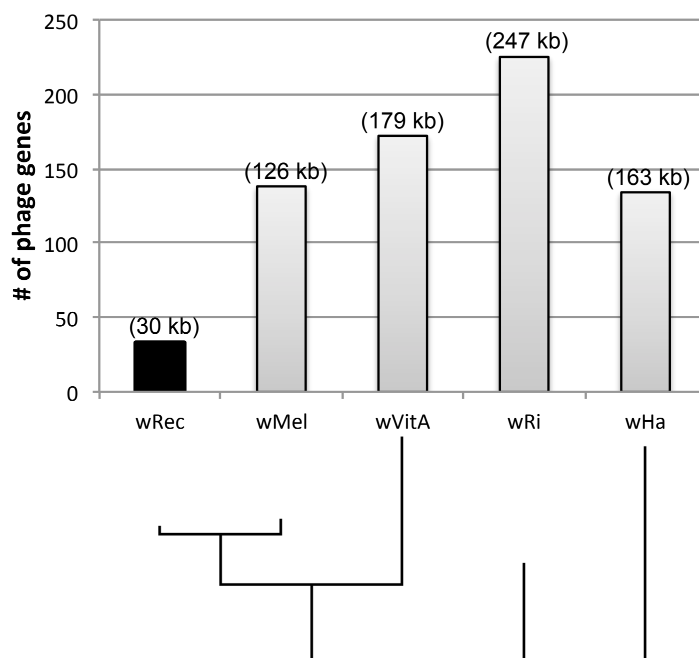


Figure IV-4. Number of phage genes in *w*Rec and its relatives. The total number of prophage, phage-associated, and WO-like island genes in each *Wolbachia* genome is plotted above a Bayesian phylogeny of their MLST genes. The approximate total length of phage genes in each genome is noted above each bar.

Discussion

Divergence and genome reduction in wRec

Genome analysis revealed that *w*Mel and *w*Rec are very closely related with an average of 99.7% nucleotide identity in coding regions shared by the two strains. We estimate that *w*Rec and *w*Mel diverged around 350,000 years ago. This estimate raises an interesting biogeographical question: how could *Wolbachia* have been transferred at this time between the widely allopatric Nearctic *D. recens* and Afrotropical *D. melanogaster*? Perhaps a widespread *Drosophila*-generalist parasitoid played a role in vectoring this endosymbiont between host species, as parasitoid wasps have been previously

demonstrated as vectors for *Wolbachia* transfer (Heath et al., 1999, Vavre et al., 2009). Molecular evidence suggests that the most recent *Wolbachia* sweep within *D. recens* occurred only 50,000 years ago, while *D. subquinaria* split from *D. recens* an estimated 600,000 years ago (Shoemaker et al., 1999). Thus, the divergence of *w*Mel and *w*Rec from their last common ancestor likely predated the most recent genetic sweep of *D. recens*, and *w*Rec infected *D. recens* after its incipient divergence from *D. subquinaria* (Werren and Jaenike, 1995, Shoemaker et al., 1999). Remarkably, these results suggest that *w*Rec may have contributed to reproductive isolation between these two species *prior* to the last glacial period 110,000 – 12,000 years ago, when their ranges are thought to have been allopatric (Jaenike et al., 2006). However, we note caution in interpreting the estimated divergence times as variability in mutation rates between bacterial lineages can skew the estimates.

The four *w*Rec genes evolving under positive selection are of particular interest as they may be potential mediators of *Wolbachia*-host interactions (Table 2). Indeed, *wsp* is known to be involved in pathogenicity and host interaction (Uday and Puttaraju, 2012) while OTU-like cysteine proteases have deubiquitinase activity facilitating the pathogenicity of intracellular pathogens and viruses (Furtado et al., 2013, Makarova et al., 2000). Although the function of the hypothetical proteins is unknown, the presence of transmembrane (TM) domains suggests interaction with the bacterial membrane and potentially its *Drosophila* host. Additionally, it has previously been speculated that the elevated rate of mitochondrial DNA evolution in *D. recens* was due to hitchhiking in association with a series of positive selection events in its resident *Wolbachia* (Shoemaker et al., 2004).

The major difference between the *w*Mel and *w*Rec genomes was the incipient genome reduction of WO prophage regions. Remaining phage WO genes in *w*Rec were often bordered or interrupted by transposases, suggesting that transposase activity may have been involved in the removal and degradation of major portions of WO genomes. Over 100kb of genetic material, consisting mostly of phage-related genes, has likely been lost in *w*Rec. Unlike the prophages found in *w*Mel (Wu et al., 2004), all of *w*Rec's WO regions lacked the head genes thought to be necessary for mature virion formation (Metcalf and Bordenstein, 2012), including the *orf7* minor capsid protein used in previous PCR tests for WO (Bordenstein and Wernegreen, 2004). The lack of such head genes suggests that *w*Rec has lost the capacity to serve as a source of WO phage to infect other strains of *Wolbachia*. Future PCR screens may benefit from inclusion of more than one primer set, perhaps adding primers for a gene from the host adsorption module, which is highly conserved across WO prophages. However, it must be cautioned that the presence of multiple and variable degenerate WO haplotypes makes it impossible for any set of one or two primer pairs to detect all haplotypes.

Meanwhile, only 2,009 SNPs were present between the *w*Mel and *w*Rec genomes, indicating that gene deletion has been heavily favored over mutation. Such genome reduction is common in obligate intracellular bacteria, where many genes are expendable due to relaxed selection and there is limited contact with novel gene pools (Casadevall, 2008). Given the predatory nature of intact WO phages (Metcalf and Bordenstein, 2012, Bordenstein et al., 2006, Sanogo and Dobson, 2006), it may have been evolutionarily advantageous for *w*Rec to eliminate the genes required for active phage production. It has been noted from TEM observations and quantitative studies that WO phage can lyse

Wolbachia, resulting in an inverse correlation between bacterial and phage densities. Furthermore, because reproductive manipulations are dependent on a critical density of *Wolbachia*, high phage activity correlates with low expression of CI (Bordenstein et al., 2006). Since *w*Rec exhibits high levels of CI in *D. recens* (Werren and Jaenike, 1995), while *w*Mel shows lower levels of CI in *D. melanogaster* (Yamada et al., 2007), it is possible that *w*Rec experiences a higher selective pressure to suppress phage, preserve high bacterial densities, and maintain compatibility with its host's mating population. Thus, this interaction could be one possible explanation for the major loss of phage genes in *w*Rec that are preserved in *w*Mel.

Although *Wolbachia* has many more repetitive and mobile elements than most obligate intracellular bacteria (Bordenstein and Reznikoff, 2005) and frequently switches hosts on an evolutionary timescale (Vavre et al., 1999), it is worthwhile to note that there were only two genes in *w*Rec that were not present in *w*Mel. It is possible that these genes were lost in *w*Mel after divergence from its last common ancestor with *w*Rec.

The phage WO hypothesis to explain reproductive parasitism

Because the Octomom region was completely absent in *w*Rec, it is unclear whether *w*Rec lost these genes after diverging from *w*Mel, or whether the genes were acquired by *w*Mel after divergence with their last common ancestor. Given that Octomom is not widespread in supergroup A *Wolbachia*, the latter possibility is likely. Moreover, although the function of Octomom in reproductive parasitism is unknown, it seems reasonable to conclude that the Octomom region is not needed for reproductive manipulations, as it is completely absent from *w*Rec. Additionally, given the association

of Octomom with increased *Wolbachia* virulence, proliferation, and host viral protection (Chrostek et al., 2013), we would predict that *w*Rec would not possess these phenotypes, and may be a useful strain for confirming these associations.

It is intriguing that some WO genes are conserved in *w*Rec while others were lost. One explanation for their preservation in *w*Rec is that the remaining genes improve *Wolbachia* fitness. Indeed, prophage sequences code for advantageous virulence factors in a wide array of bacterial species (Brussow et al., 2004). Because previous PCR surveys suggested *w*Rec did not possess phage WO, speculation that WO may be involved in *Wolbachia* reproductive manipulations has been largely disregarded (Bordenstein and Wernegreen, 2004). However, our sequencing shows that although the phage genomes are not complete, *w*Rec contains many phage-related genes including some that could be involved in CI and/or male-killing. These include at least four ankyrin repeat proteins, whose repetitive domain has been long thought to facilitate *Wolbachia*-eukaryote interaction (Iturbe-Ormaetxe et al., 2005, Siozios et al., 2013). Additionally, several WO genes in *w*Rec are homologs of genes recently implicated in CI. WREC_0560 is a transcriptional regulator with 88.3% identity at the amino acid level to *wtrM* in *w*PipMol, which increases expression of an important regulator of meiosis in *Culex* mosquitos and is postulated to be a component of the molecular mechanisms of CI (Pinto et al., 2013). WREC_0566-WREC_0568 meanwhile, are homologous to WP_0282 and WP_0283, two genes in *w*Pip that have been implicated in CI due to presence in the proteome of *Wolbachia*-infected, fertilized mosquito spermathecae, along with their pattern of presence/absence in CI and non-CI strains (Beckmann and Fallon, 2013). Although the *w*Rec homolog of WP_0283 has been truncated by 427bp, it has 99.8% nucleotide

identity to the gene in *w*Mel (WD_0632) and an alternative reading frame enables the transcription of the remaining nucleotides in the same frame as the C-terminus of the homolog in *w*Mel. Whether any of these WO genes are actually involved in *Wolbachia* host manipulations remains unclear, especially since it is unknown whether the remnants of phage WO are transcribed by *w*Rec. However, the fact that these prophage regions are conserved suggests that they may have a role to play in the biology of *Wolbachia*.

WO host adsorption genes

In addition to preservation of some potential reproductive manipulation mediators, prophage WO genes WREC_0263-WREC_0269 contain an intact host adsorption module that includes baseplate genes thought to be involved in the binding of WO to its bacterial host and insertion of phage DNA. Indeed, this host adsorption module is nearly universal in WO prophage, with very few degenerate phage haplotypes lacking these genes (Kent et al., 2011a). Many intracellular bacteria, including *Wolbachia* (Rances et al., 2008, Pichon et al., 2009), possess a type IV secretion system that secretes effectors into the host as a common strategy to subvert host-cell functions (Voth et al., 2012). A number of Gram-negative bacteria also possess a phage-like type VI secretion system (Coulthurst, 2013); these include several obligatory intracellular bacterial pathogens, such as *Anaplasma* and *Ehrlichia* (Rikihisa and Lin, 2010). Structural analyses have shown this type of secretion system bears a remarkable resemblance to the spike protein of phages (Silverman et al., 2012). Given these similarities, and the fact that the WO host adsorption module is almost universally present in sequenced arthropod *Wolbachia* (Kent et al., 2011a), it is possible that *Wolbachia* may be using these genes to

facilitate host-microbe interactions, as a way to inject CI factors, genes, or other host manipulation particles into its host.

Future Studies

The discovery of phage elements in *w*Rec opens up new questions. Additional experiments will be needed to determine whether any conserved phage genes are involved in *Wolbachia* manipulations of its host biology. In addition, we have seen that using single gene markers of phage WO is not diagnostic of its absence. Thus, unsequenced *Wolbachia* strains that were thought to be phage-free by PCR assays need reevaluation. Another question that remains is how a single *Wolbachia* causes multipotent reproductive manipulations in different host backgrounds. The availability of genomic sequence for a multipotent *Wolbachia* strain will enable future transcriptomic and proteomic studies that could elucidate the genes involved in switching reproductive phenotypes.

CHAPTER V. *WOLBACHIA*-INDUCED CYTOPLASMIC INCOMPATIBILITY MAY BE CAUSED BY PHAGE-ENCODED VIRULENCE FACTORS[§]

Abstract

Bacteria are capable of complex manipulations of their hosts, especially among maternally inherited endosymbionts that can skew sex ratios, sex determination, or fitness to favor infected females. One particularly widespread manipulation is the phenomenon of cytoplasmic incompatibility (CI), a crossing incompatibility induced by *Wolbachia pipientis* in which the offspring of uninfected females die when mated with infected males; infected females can rescue this defect to produce viable and infected offspring. Despite decades of investigation and important applications to insect speciation and pest control, the mechanisms by which *Wolbachia* induce CI are unknown. Here, we use a set of four criteria to predict potential *Wolbachia* genes that might be mediators of CI and select six genes for functional testing. Interestingly, five of these genes are encoded by prophage WO and exhibit differential tissue expression in infected ovaries and testes. Two of these genes cause a dramatic drop in hatch rates when transgenic males are mated to wild type females. Consistent with a CI effector, this defect is completely rescued by infected females. Although more work must be done, these experiments are the first to recapitulate CI using *Wolbachia* transgenes and make exciting inroads into understanding the mechanisms of endosymbiont-mediated reproductive manipulation. We conclude with a discussion on how CI genes could be deployed to control insect vectors of disease.

[§] Daniel LePage and Seth R. Bordenstein contributed to the authorship of this chapter.

Introduction

At least half a dozen genera of arthropod endosymbionts evolved mechanisms to manipulate the reproduction of their hosts to favor the success of infected females (Hurst and Jiggins, 2000, Cordaux et al., 2011, Engelstadter and Hurst, 2009, Engelstadter and Telschow, 2009). The most common of these manipulations is cytoplasmic incompatibility (CI), a form of embryonic lethality that occurs between infected males and females that are not infected (unidirectional CI) or infected with a different bacterial strain than that in the male (bidirectional CI). Females infected with the same strain however, can rescue this defect and produce viable, infected offspring when crossed with either uninfected or infected males, thereby favoring the spread of this maternally-inherited bacterium in a host population with variable infection status (Engelstadter and Telschow, 2009). This makes CI a key player in driving speciation events, causing reproductive isolation between populations infected with different strains or reinforcing speciation of infected and uninfected species through behavioral isolation (Jaenike et al., 2006, Bordenstein et al., 2001, Miller et al., 2010). CI has important applications in disease vector control (LePage and Bordenstein, 2013), and is currently being used in field trials to drive the spread of Dengue-resistant mosquitoes in wild populations through the release of *Wolbachia*-infected females (Walker et al., 2011, Bull and Turelli, 2013), and as a biological control mechanism to depress mosquito populations by releasing *Wolbachia*-infected males incompatible with wild females (O'Connor et al., 2012). Despite decades of research (Yen and Barr, 1971, Breeuwer and Werren, 1993, Bordenstein et al., 2001, Poinsot et al., 2003, Serbus et al., 2008), the causative effectors of cytoplasmic incompatibility are still unknown. However, a significant body of work

has elucidated the host cytological defects that occur in afflicted embryos, strain and host variation in CI expression, and gene expression in infected gonads. This combinatorial information enables a structured approach to prediction of potential CI mediators. We sought candidate CI effectors based on the following four criteria:

Criterion #1: Presence/absence in CI strains

While most *Wolbachia* strains infecting arthropods cause CI, male-killing, or other parasitic phenotypes facilitating *Wolbachia* spread, strain *w*Au does not induce any known form of reproductive manipulation (Hoffmann et al., 1996). Although *w*Au has not been sequenced, a microarray has been performed in search of absent or divergent genes compared to closely related CI-inducing relatives (Ishmael et al., 2009). Meanwhile, *Wolbachia* strains infecting nematodes, such as *w*Bm, are mutualists rather than parasitic (Fenn and Blaxter, 2004), and *w*Bm has a sequenced genome that is quite different from the strains that cause CI (Casiraghi et al., 2005). A number of CI-inducing *Wolbachia* strains have also been sequenced (Klasson et al., 2008, Wu et al., 2004, Klasson et al., 2009b). We recently sequenced another strain, *w*Rec (Metcalf et al., 2014), which causes both CI and male-killing depending on its host (Jaenike, 2007), and possesses a reduced gene set compared to its relatives. Thus, two datasets are available for a comparative genomics analysis between CI and non-CI inducing strains. Criterion 1a includes all genes absent or divergent in *w*Au but present in *w*Ri or *w*Sim as detected by microarray (Ishmael et al., 2009) (Appendix Table A1). Criterion 1b includes all genes present in the core genome of CI-inducing arthropod strains (*w*Mel, *w*Ri, *w*Rec, and *w*Pip), but absent from the pan-genome of *w*Bm (Appendix Table A2).

Criterion #2: Expression in gonads

Since CI is caused by a modification of sperm and can be rescued by a *Wolbachia*-infected egg, CI effectors must be present in at least one of the arthropod gamete-producing tissues, namely ovaries and testes. However, caution should be used with this criterion, as rare *Wolbachia* transcripts or proteins may be involved and could be missed with omics approaches. To date, two proteomic analyses of *Wolbachia*-infected reproductive tissues in mosquitos have been published (Beckmann et al., 2013, Beckmann and Fallon, 2013), and a third experiment examining the proteome of infected *Drosophila simulans* flies has been performed (Daniel LePage, unpublished data). Data at the level of transcription is also available, in the form of RNA-seq from *w*VitA-infected *Nasonia vitripennis* ovaries (Lisa Funkhouser-Jones, unpublished data). Criterion 2a and 2b encompasses all *Wolbachia* gene products present in reproductive tissues at the level of protein and RNA, respectively (Appendix Table A3, A4).

Criterion #3: Secretion

CI effectors must presumably access host cells to manipulate the arthropod genome or cell cycle. *Wolbachia* contains a type IV secretion system (T4SS) that likely mediates export of CI effectors into the host cytoplasm (Pichon et al., 2009). Although a conserved secretion signal for type IV effectors does not appear to exist, bioinformatic analyses can predict these effectors with some success, using a variety of factors such as homology to known effectors, homology to eukaryotic proteins, cellular localization signals, and the amino acid composition in specific protein regions (Meyer et al., 2013).

Criterion 3 comprises all genes with a high likelihood score of being secreted (Appendix Table A5). Again, this criterion must be used with caution, as the prediction algorithm is far more accurate on γ-proteobacteria compared to α-proteobacteria like *Wolbachia*. Additionally, it is possible that the CI effectors are secreted by an alternative means, independent of the T4SS.

Criterion #4: Previous candidate testing

Thirteen *Wolbachia* genes have already been tested for their involvement in CI through transgenic expression in *D. melanogaster* (Yamada et al., 2011). These candidates consisted almost entirely of ankyrin repeat proteins, a domain long speculated to be involved in *Wolbachia*-mediated reproductive manipulations (Iturbe-Ormaetxe et al., 2005). However, despite rigorous testing, none of these candidates could recapitulate or rescue CI, either alone or in combination with *w*Au. Thus, these genes were excluded as candidate CI effectors (Appendix Table A6).

Finalizing candidates: function, toxin-antitoxin systems, and bidirectional incompatibility

Final CI candidates were selected based on the above criteria and further narrowed based on putative gene functions, toxin-antitoxin system prediction, and patterns of bidirectional incompatibility. Despite a lack of understanding of the bacterial effectors of CI, much is known about its pathophysiology in insects. Cytological defects in CI embryos include delayed breakdown of the nuclear envelope, abnormal deposition of maternal histones, incomplete paternal DNA replication, and failure of chromosome condensation, leading to disruption of the first mitosis and embryonic lethality

(Landmann et al., 2009, Serbus et al., 2008, Tram and Sullivan, 2002). Thus, *Wolbachia* proteins that may be involved in chromosomal interactions, DNA or histone modification, regulation of the cell cycle, or global alterations of transcription or post-translational modifications would all be of particular interest as CI effector candidates.

Type II toxin-antitoxin (TA) systems can be found in most bacterial genomes and are generally identified by the presence of two genes, a stable toxin and labile antitoxin, organized within an operon (Yamaguchi et al., 2011). With few exceptions, the genes often overlap and the larger toxin is located downstream from the smaller antitoxin. The toxin is rendered inactive during dormant conditions by protein binding to its cognate antitoxin and, upon times of stress, the antitoxin is degraded rendering the toxin free to induce cell stasis or death. TA systems have been linked with stress response, programmed cell death, persistence and biofilm formation. A few TA loci, including *relBE*, stabilize large mobile genetic elements and limit the amount of gene loss associated with reductive evolution (Szekeres et al., 2007). TA systems were initially considered to be limited to free-living prokaryotes (Pandey and Gerdes, 2005), though the growing number of sequenced obligate intracellular symbionts containing TA loci contradicts this hypothesis (Yamaguchi et al., 2011). A leading model for CI is the lock and key mechanism (Poinsot et al., 2003), sometimes described as a toxin-antitoxin system, though not in the classical sense. Recently, it has been shown that a Rickettsia TA system is toxic to its arthropod host (Audoly et al., 2011), suggesting that CI may be caused by a similar mechanism. Thus, any CI candidates that may be part of a TA system are of particular interest.

A number of studies have demonstrated patterns of bidirectional incompatibility between *Wolbachia* strains in which CI induced in males infected by one strain cannot be rescued by females infected with the other strain and vice versa. Although not all of these studies have been well controlled, there are several that could be used to test whether the relatedness of CI effector candidates found here and their homologs in other strains can explain bidirectional CI (Zabalou et al., 2008, Pinto et al., 2013). These patterns of relatedness were used to further support and refine our list of candidates.

Materials and Methods

Candidate selection

For criterion 1a, microarray data were obtained from the original authors (Ishmael et al., 2009) and genes that were present in CI-inducing strains *w*Ri and *w*Sim but absent or divergent in *w*Au were selected. For criterion 1b, MicroScope (Vallenet et al., 2009) was used to select the set of genes comprising the core genomes of CI-inducing strains *w*Pip, *w*Mel, *w*Rec, and *w*Ri, but absent from the pan-genome of the mutualistic *w*Bm, using cutoffs of 50% amino acid identity and 80% alignment coverage. Proteome datasets were obtained from *w*Pip-infected *Culex pipiens* gonads (Beckmann et al., 2013), *w*Pip-infected and fertilized *C. pipiens* spermathecae (Beckmann and Fallon, 2013), and *w*Ri-infected *D. simulans* sperm (Daniel LePage, unpublished data) for criterion 2a. The 25 genes with the greatest number of *Wolbachia* RNA-seq reads in a transcriptome of *w*VitA-infected *Nasonia vitripennis* ovaries (Lisa Funkhouser-Jones, unpublished data) were selected for criterion 2b. The closest homologs in strain *w*Mel were found for each

selected gene in the proteome and transcriptome data sets using blastp in Geneious Pro v5.5.6 (Kearse et al., 2012). The Searching Algorithm for Type IV Effector proteins (S4TE) was used to detect genes in the *w*Mel genome with a high probability of being secreted (Meyer et al., 2013), and any genes passing six or more secretion tests out of the 13 tests available were selected for criterion 3. Genes that were previously tested in transgenic *D. melanogaster* (Yamada et al., 2011) were excluded and a final list of candidates was determined based on predicted gene function, relatedness to homologs in strains exhibiting bidirectional incompatibility, and the number of above criteria satisfied. Homologs of effector candidates were obtained with BLASTp of the nr database and aligned with MUSCLE (Edgar, 2004) using Geneious 5.5.9 (Kearse et al., 2012). Identity matrices were used to search for patterns of relatedness that may explain bidirectional incompatibility relationships.

Candidate gene expression

Expression of CI candidates was tested with RT-qPCR on dissected ovaries from one-day-old *w*Mel-infected *D. melanogaster* (W^{1118}) virgin females and dissected testes from one-day-old and seven-day-old virgin males. Samples contained either 10 pairs of ovaries or 20 pairs of testes. 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). Candidate gene expression was compared to *groEL* for normalization to *Wolbachia* housekeeping gene expression or *gpW* to control for phage WO activity. Primers are listed in Table V-1.

Table V-1. Primers used in RT-qPCR.

Gene	Primer (5'->3')
gpW-F	CTACAACCTCATCGAAGCGAATCT
gpW-R	CTGCAGAAGCTTTGGAAAAATGGG
GroEL-F	CAACCTTTACTTCCTATTCTTG
GroEL-R	CTAAAGTGCTTAATGCTTCACCTTC
WD0034-F	GGAAGAAACTTGCACACCACTTAC
WD0034-R	TGCTCTCCGACCATCTGGATATTT
WD0508-F	TAGAGATCTAGCTTGCGGACAAGA
WD0508-R	TCCTTAACTAAACCCTTTGCCACC
WD0625-F	GAGCCATCAGAAGAAGATCAAGCA
WD0625-R	TTCTCGAAAGCTGAAATAGCCTCC
WD0626-F	AATTGGCCTCTCTGCTAATGAGTG
WD0626-R	CACGTCCTTGCTCATAGTTGCTTA
WD0631-F	TGTGGTAGGGAAGGAAAGAGGAAA
WD0631-R	ATTCCAAGGACCATCACCTACAGA
WD0632-F	TGCGAGAGATTAGAGGGCAAAATC
WD0632-R	CCTAAGAAGGCTAATCTCAGACGC

Transgenic Drosophila

Each CI candidate gene was cloned into the pTIGER plasmid for transformation and expression in *D. melanogaster* (Ferguson et al., 2012). pTIGER was designed for targeted integration into the *D. melanogaster* genome using PhiC31 integrase (Groth et al., 2004) and tissue-specific, inducible expression through the Gal4-UAS system (Southall et al., 2008). Cloning was performed using standard molecular biology

techniques and plasmids were purified and sequence-confirmed 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. Isogenic, homozygous lines were maintained when possible, or isogenic heterozygous flies were maintained when homozygous transgenics were inviable (WD0625/CyO).

CI and rescue crosses

Parental flies for hatch rate experiments were created by crossing *nanos*-Gal4 virgin females (*w*Mel infected or uninfected) with either wild type or UAS-candidate gene-expressing males. Hatch rate assays were performed using a grape juice/agar media in 30mm plates for egg laying. For each cross, 32-48 individual crosses of one male and one female were set up in separate mating chambers with individual grape juice plates. A minimal amount of a 1:2 dry yeast and water mix was added to each plate; the parents were allowed to mate for 16 hours, and the grape juice plates were discarded. Fresh plates were then used for 24 hours, removed, and the number of eggs laid for each cross was counted. Parental flies were also discarded at this time. The number of unhatched eggs was counted again at 36 hours after the plates had been removed to determine hatch rates. Larvae were also removed from 10 individual crosses in selected experiments to test the candidate genes' capabilities to alter host sex ratios.

Results and Discussion

Selection of candidate genes for further evaluation

Filtering the complete *w*Mel genome through criteria 1-4 selected 232 unique genes (criterion 1a: 60 genes, criterion 1b: 128 genes, criterion 2a: 7 genes, criterion 2b: 25 genes, criterion 3: 30 genes, criterion 4: 12 genes excluded), only 14 of which matched more than one inclusion criterion without being excluded by previous testing (Table V-2). A complete list of genes matching each CI effector criterion is available in Appendix A. Interestingly, nine of the genes matching multiple criteria were located in prophage regions (Kent et al., 2011a) or the phage-like island Octomom (Chrostek et al., 2013).

Table V-2. *w*Mel genes matching two or more of CI candidate criteria #1-3 but not criterion #4.

Locus	Function	Criteria
WD0035	ankyrin repeat-containing protein	1b, 3
WD0147	tetratricopeptide repeat family protein	2b, 3
WD0255	transcriptional regulator	1a, 1b
WD0286	ankyrin repeat-containing prophage LambdaW1	1b, 3
WD0292	ankyrin repeat protein	2b, 3
WD0424	hypothetical protein	1a, 3
WD0508	transcriptional regulator	1a, 1b
WD0623	transcriptional regulator	1a, 1b
WD0625	DUF2466 nuclease	1a, 1b
WD0626	transcriptional regulator	1a, 1b
WD0631	hypothetical protein	1a, 1b, 2a, 2b
WD0632	SUMO protease	1a, 1b, 2b
WD0696	hypothetical protein	1b, 3
WD1063	wsp surface protein	2a, 2b

Genes in phage regions or phage-like islands are bolded. Genes selected as candidates for experimental testing are highlighted. A sixth gene, WD0034, was also selected for testing despite only matching one criterion, due to its suspected function in RNA silencing (see text).

The list of 14 CI candidates was further narrowed based on putative gene function, toxin-antitoxin system prediction, and bidirectional incompatibility patterns to select a final list of genes for functional analysis. WD0631 and WD0632 were of particular interest due to the number of criteria satisfied, as well as their potential functions. Although no known domains are present in WD0631, there is evidence that it is transcribed as an operon with WD0632 (Beckmann and Fallon, 2013), which contains a small ubiquitin-related modifier (SUMO)-associated Ulp1 protease domain in addition to a domain of unknown function. The presence of a SUMO protease in *Wolbachia* is notable because SUMOylation is an exclusively eukaryotic posttranslational protein modification not found in bacteria, although it may be manipulated by microbes as a mechanism of pathogenesis (Wimmer et al., 2012). Ulp1 SUMO proteases are crucial enzymes in eukaryotes, and functional mutations can lead to cell cycle arrest (Li and Hochstrasser, 1999). Additionally, SUMOylation is very important for regulation of chromatin structure, and is also involved in DNA methylation (Cubenas-Potts and Matunis, 2013), two phenomena that can be altered during *Wolbachia*-induced CI (Landmann et al., 2009, Ye et al., 2013). WD0631/WD0632 also satisfy conditions for a TA system, and WD0631 patterns of relatedness correlate with bidirectional incompatibility relationships (see below). Given this information, WD0631 and WD0632 were selected for further analysis.

Transcriptional regulation is a known mechanism of bacterial effectors mediating pathology in numerous bacterial species (Dean, 2011). For example, a family of effectors in *Xanthomonas* localizes to the host nucleus, mimics eukaryotic transcription factors, and reprograms plant cells to facilitate pathogenesis (Kay et al., 2007). Notably,

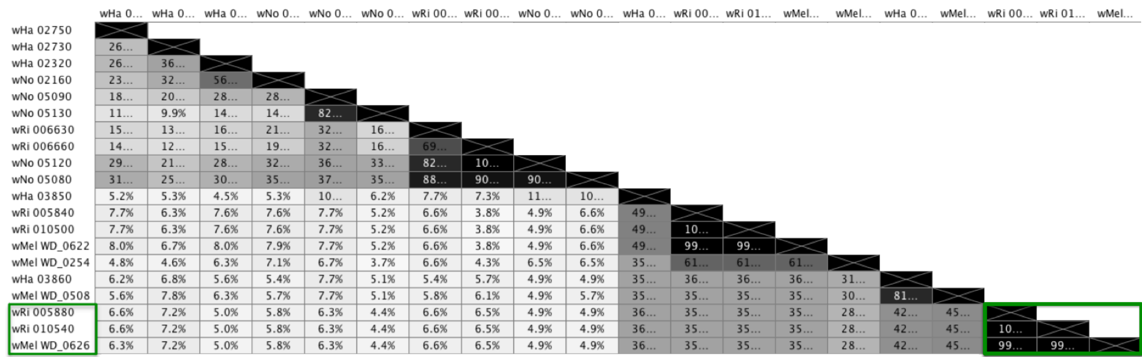
transformation of mosquitoes with a WO-associated *Wolbachia* transcriptional regulator from *w*Pip (homolog in wMel is WD0626) upregulated expression of a gene involved in regulation of meiotic division (Pinto et al., 2013). A number of studies have noted changes in host transcription during *Wolbachia* infection (Pinto et al., 2013, Zheng et al., 2011, Xi et al., 2008a, Clark et al., 2006). Thus, *Wolbachia* transcriptional regulators could potentially alter transcription of key host factors involved in meiosis as a means of inducing CI, and two of these genes, WD0508 and WD0626 were selected for further characterization. WD0626 is of additional interest as its patterns of relatedness also correlate with bidirectional incompatibility (see below), and may be part of a TA system with WD0625. WD0507/WD0508 are homologous to WD0625/WD0626, but WD0507 is disrupted and likely nonfunctional. Nonetheless, WD0508, the putative antitoxin, could still be active.

WD0625 contains a DUF2466 (domain of unknown function) domain formerly annotated as the DNA repair protein RadC, but has since been renamed (Attaiech et al., 2008) and implicated as a putative nuclease (Iyer et al., 2011). Degradation of host mRNA is a common strategy used by viruses to manipulate their hosts (Gaglia et al., 2012), and paternal sperm RNA may play an important role in early embryo development (Jenkins and Carrell, 2012). Thus, WD0625 was also added to the list for functional tests. Finally, a sixth gene, WD0034 was selected for further studies based on absence in *w*Au (criterion 1a). Despite lacking a match to other criteria, WD0034 is of interest because it contains three PAZ (Piwi, Argonaut, and Zwille) domains, which are common in proteins involved in RNA silencing (Chen and Varani, 2005). Several recent studies have found

perturbation of microRNAs in *Wolbachia*-infected cells (Osei-Amo et al., 2012, Mayoral et al., 2014, Zhang et al., 2013), making WD0034 a potential mediator of these changes.

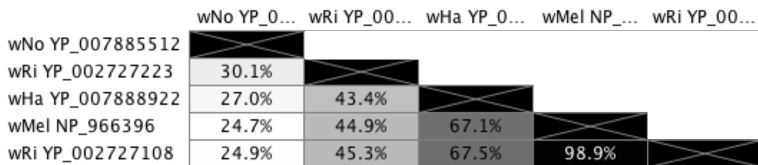
Bidirectional compatibility

To check for patterns of relatedness in CI effector candidates that could explain bidirectional incompatibility, we obtained homologs of effector candidate genes from sequenced *Wolbachia* genomes having well-defined bidirectional incompatibility patterns and organized them into families based on similarity. Interestingly, strains with very close homologs of WD0626 (Figure V-1) and WD0631 (Figure V-2) were compatible, while strains with more distantly related homologs were not. WD0626 is nearly identical to its closest homolog in *w*Ri, but only divergent homologs are present in *w*No and *w*Ha. Based upon this, if WD0626 is a major mediator of CI, we would predict that *w*Mel and *w*Ri would be compatible, but neither strain would be compatible with *w*No or *w*Ha. Although *w*Mel's compatibility with wNo and wHa has not been tested, the remaining combinations are as predicted (Zabalou et al., 2008). WD0631 has a similar pattern, with nearly identical copies in wMel and wRi and divergent homologs in wNo and wHa, with the same predicted outcome (Figure V-2).



	wHa 0...	wHa 0...	wHa 0...	wNo 0...	wNo 0...	wNo 0...	wRi 00...	wRi 00...	wNo 0...	wNo 0...	wHa 0...	wRi 00...	wRi 01...	wMel...	wMel...	wHa 0...	wMel...	wRi 00...	wRi 01...	wMel...
wHa 02750																				
wHa 02730	26...																			
wHa 02320	26...	36...																		
wNo 02160	23...	32...	56...																	
wNo 05090	18...	20...	28...	28...																
wNo 05130	11...	9.9%	14...	14...	82...															
wRi 006630	15...	13...	16...	21...	32...	16...														
wRi 006660	14...	12...	15...	19...	32...	16...	69...													
wNo 05120	29...	21...	28...	32...	36...	33...	82...	10...												
wNo 05080	31...	25...	30...	35...	37...	35...	88...	90...	90...											
wHa 03850	5.2%	5.3%	4.5%	5.3%	10...	6.2%	7.7%	7.3%	11...	10...										
wRi 005840	7.7%	6.3%	7.6%	7.6%	7.7%	5.2%	6.6%	3.8%	4.9%	6.6%	49...									
wRi 010500	7.7%	6.3%	7.6%	7.6%	7.7%	5.2%	6.6%	3.8%	4.9%	6.6%	49...	10...								
wMel WD_0622	8.0%	6.7%	8.0%	7.9%	7.7%	5.2%	6.6%	3.8%	4.9%	6.6%	49...	99...	99...							
wMel WD_0254	4.8%	4.6%	6.3%	7.1%	6.7%	3.7%	6.6%	4.3%	6.5%	6.5%	35...	61...	61...	61...						
wHa 03860	6.2%	6.8%	5.6%	5.4%	7.7%	5.1%	5.4%	5.7%	4.9%	4.9%	35...	36...	36...	36...	31...					
wMel WD_0508	5.6%	7.8%	6.3%	5.7%	7.7%	5.1%	5.8%	6.1%	4.9%	5.7%	35...	35...	35...	35...	30...	81...				
wRi 005880	6.6%	7.2%	5.0%	5.8%	6.3%	4.4%	6.6%	6.5%	4.9%	4.9%	36...	35...	35...	35...	28...	42...	45...			
wRi 010540	6.6%	7.2%	5.0%	5.8%	6.3%	4.4%	6.6%	6.5%	4.9%	4.9%	36...	35...	35...	35...	28...	42...	45...	10...		
wMel WD_0626	6.3%	7.2%	5.0%	5.8%	6.3%	4.4%	6.6%	6.5%	4.9%	4.9%	36...	35...	35...	35...	28...	42...	45...	99...	99...	

Figure V-1. Identity matrix of WD0626 homologs in wMel and its relatives. WD0626 and its closest homologs are boxed in green. Gene pair similarity is indicated by shading, with darker shading for pairs with the highest similarity.



	wNo YP_0...	wRi YP_00...	wHa YP_0...	wMel NP_...	wRi YP_00...
wNo YP_007885512					
wRi YP_002727223	30.1%				
wHa YP_007888922	27.0%	43.4%			
wMel NP_966396	24.7%	44.9%	67.1%		
wRi YP_002727108	24.9%	45.3%	67.5%	98.9%	

Figure V-2. Identity matrix for WD0631 homologs in relatives of wMel. Gene pair similarity is indicated by shading, with darker shading for pairs with the highest similarity.

A second well-established bidirectional incompatibility that could be tested with these candidates is that between strains of *Culex* mosquito *Wolbachia* - *w*Pip Mol, *w*Pip Pel, and *w*Pip JHB. Pel and JHB are compatible, while Mol is incompatible with both Pel and JHB. Thus, any differences between Mol and the other two could be involved in bidirectional CI. Interestingly, only three genes are present in Mol but absent in Pel, and 14 genes are present in Pel but absent in Mol (Pinto et al., 2013). Two transcriptional regulators homologous to WD0626 are among these genes, suggesting that the WD0626 family of transcriptional regulators might be involved in bidirectional CI. WD0631

homologs are present in both wPip Mol and wPip Pel, but the homolog in Mol is much shorter than the homolog in Pel (172 vs. 491 aa, 90.1% identity; Figure V-3).

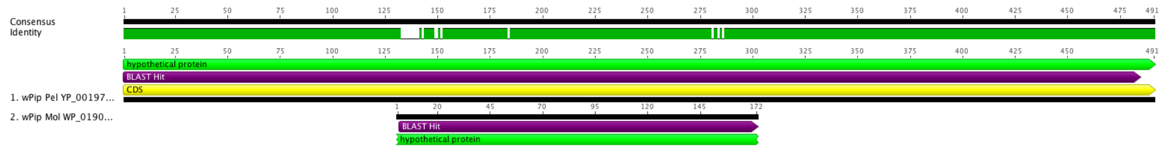


Figure V-3. WD0631 homologs in wPip Pel (top) and wPip Mol (bottom). Although sharing 90.1% pairwise identity, the homolog in Mol is lacking much of the protein in Pel.

Candidate gene expression patterns

The expression of CI generally decreases as male insects age (Reynolds and Hoffmann, 2002). Thus, it is possible that the expression of the sperm-modifying factor that is the cause of CI will also decrease as males age. Additionally, the gene(s) that rescue CI must be expressed in *Wolbachia*-infected ovaries, and might be differentially expressed relative to male testes. Therefore, we examined the expression of each CI candidate gene in young (one-day-old) and old (seven-day-old) *D. melanogaster* testes, and in one-day-old ovaries (Figure V-4). Interestingly, all five phage genes, but not *Wolbachia* gene WD0034, exhibited some level of statistically significant, tissue-specific expression in a pattern that could be consistent with a CI effector.

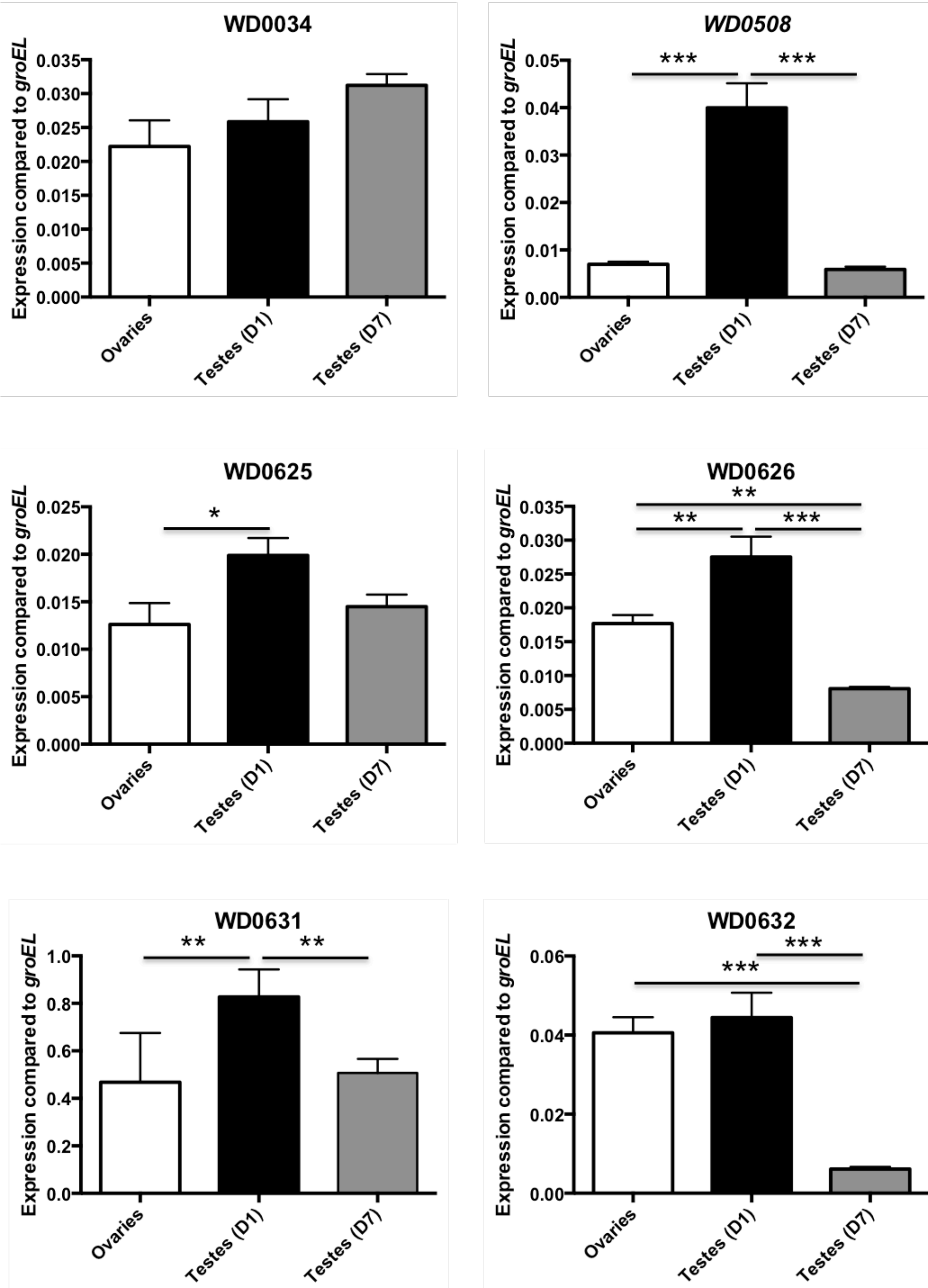


Figure V-4. Expression of CI candidate effectors in reproductive tissues. The expression of each candidate effector is shown relative to the *Wolbachia* housekeeping gene *groEL* in the ovaries of one-day-old females or testes of either one-day-old (D1) or seven-day-old (D7) males. *** P<0.001, ** P<0.01, * P<0.05 for one-way ANOVA with Tukey multiple comparisons test.

Functional tests of CI

To test whether CI effector candidates could recapitulate or rescue *Wolbachia*-induced CI, we created transgenic *D. melanogaster* expressing each candidate in the germ line and performed hatch rate assays to determine the effect of transgenes on embryo viability. Strikingly, both WD0625 and WD0626 dropped embryo viability when expressed in males, a defect that was fully rescued by *Wolbachia*-infected females (Figure V-5). Hatch rates dropped from 85% (experiment 1) or 95% (experiment 2) in wild type crosses to 56% (experiment 1) or 80% (experiment 2) when males expressing WD0625 mated with uninfected females (P<0.001 in one-way ANOVA with Tukey multiple comparison test for both experiments). Crossing these WD0625 transgenic males with *w*Mel-infected females restored hatch rates to 94% (P=0.001 vs. induced CI cross, not significant vs. wild type cross). Meanwhile, hatch rates decreased from 91% in wild type crosses down to 50% when males expressing WD0626 mated with uninfected females (P<0.001), a defect fully rescued by infected females (94% hatch rate, P<0.001 vs. induced CI cross, not significant vs. wild type cross). WD0625 is a putative nuclease, while WD0626 is annotated as a transcriptional regulator. As both genes may potentially interact with nucleic acid and both recapitulate *Wolbachia*-induced CI, these two genes were selected as our strongest candidates for more in-depth analysis in ongoing experiments further described in Chapter VII.

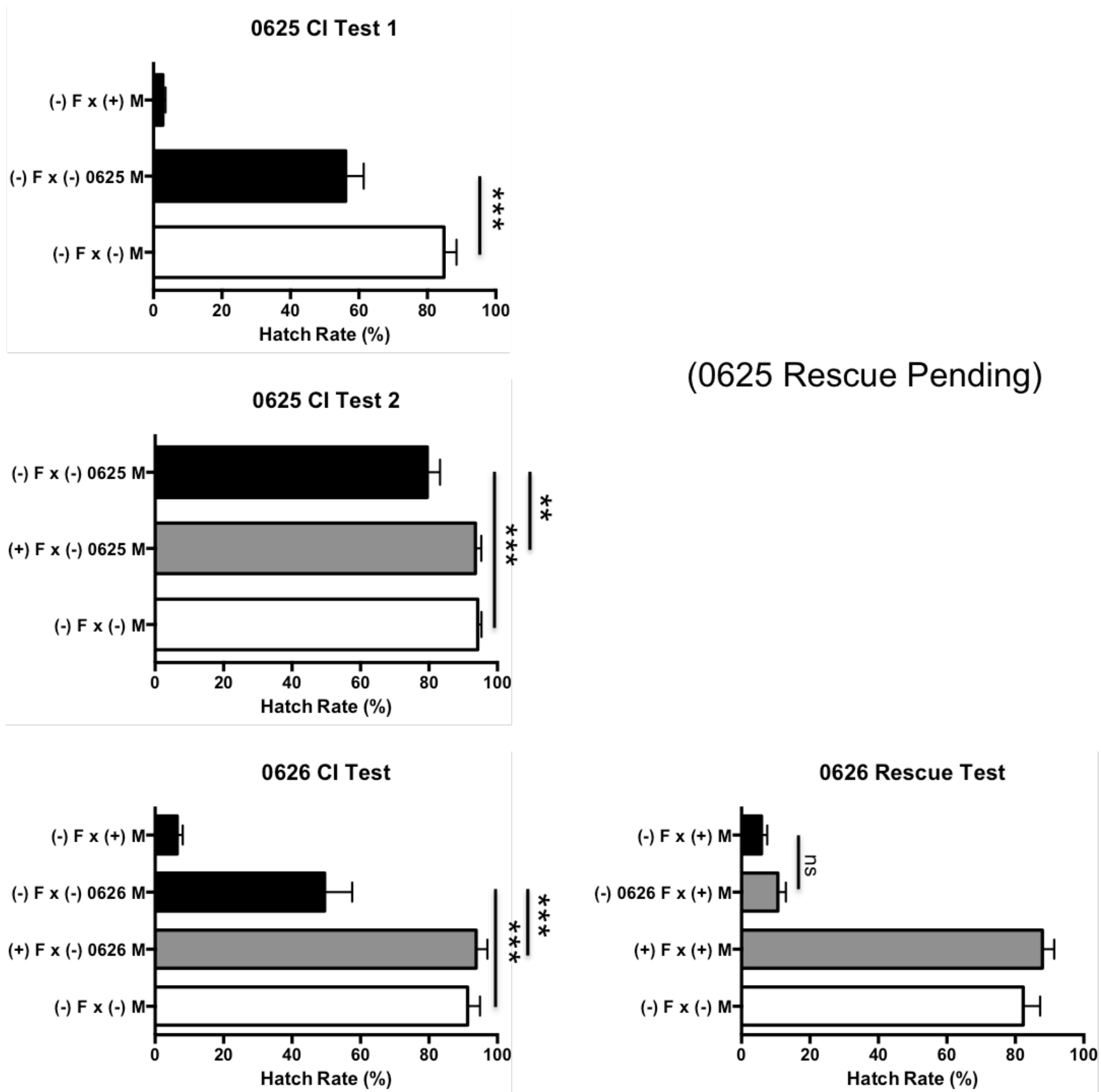


Figure V-5. Hatch rate assays for *D. melanogaster* expressing either WD0625 or WD0626. Infection status of the female (F) and male (M) of each cross is indicated as + (infected with *w*Mel) or - (uninfected). At least 32 individual crosses were performed for each assay. *** P<0.001, ** P<0.01, * P<0.05 for one-way ANOVA with Tukey multiple comparisons test. Not all pairwise comparisons are shown.

The four other CI effector candidates either induced no CI (WD0034 and WD0632), minimal CI (WD0508), or a CI that was not rescuable (WD0631), in these preliminary experiments (Figure V-6). There were no statistically significant differences

in matings between uninfected females and transgenic males expressing WD0034 or WD0632, suggesting that they are unlikely to be involved in CI. Interestingly, hatch rates did decrease from wild type rates of 96% to 87% in matings between uninfected females and WD0508-expressing males (P=0.035), and infected females significantly rescued these hatch rates to 97% (P=0.008 vs. induced CI cross, not significant vs. wild type cross). These data suggest that WD0508, which is a transcriptional regulator in the same family as WD0626, may have a small role to play in CI, but additional hatch rate assays to verify this more subtle defect are currently planned.

WD0631 on the other hand, induced a substantial drop in hatch rates from 82% in wild type crosses to only 42% in crosses between uninfected females and WD0631-expressing males (P<0.001). However, this decrease was not rescuable by *Wolbachia*-infected females (60% hatch rate, not significant vs. induced CI cross, P<0.001 vs. wild type cross). It is possible that WD0631 is toxic to embryos in a fashion independent of *Wolbachia*-induced CI, or that the transgene is expressed at too high a level for *Wolbachia* to rescue the defect. These possibilities are being investigated by a hatch rate assay including a cross in which WD0631 is expressed in uninfected females, and by qPCR to compare transgene expression levels to expression in *Wolbachia*-infected wild type flies.

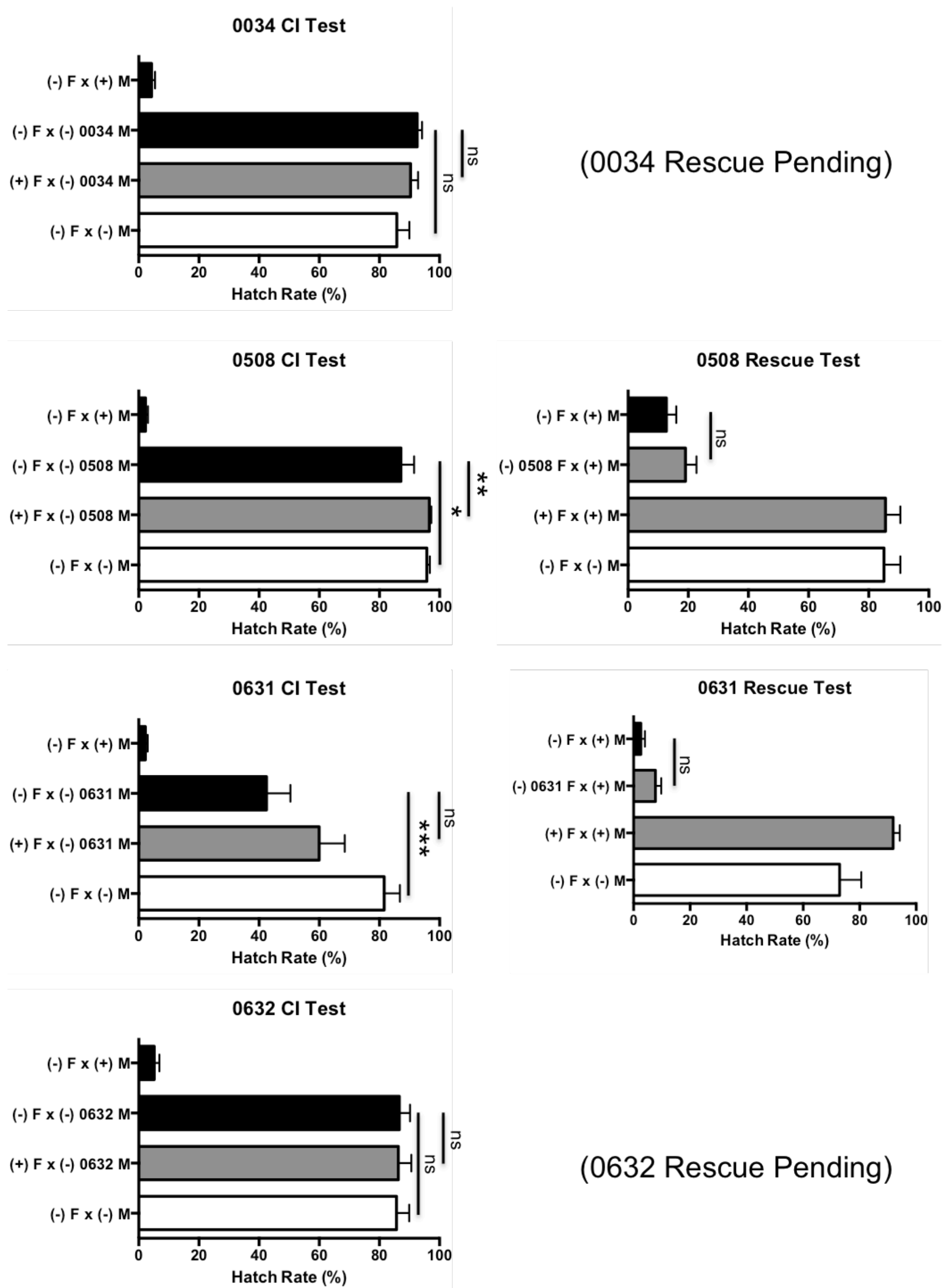


Figure V-6. Hatch rate assays for *D. melanogaster* expressing the indicated transgenes. Infection status of the female (F) and male (M) of each cross is indicated as + (infected with *w*Mel) or - (uninfected). At least 32 individual crosses were performed for each assay. *** P<0.001, ** P<0.01, * P<0.05 for one-way ANOVA with Tukey multiple comparisons test. Not all pairwise comparisons are shown.

Conclusions

To the best of our knowledge, these experiments are the first to demonstrate any modification of hatch rates caused by *Wolbachia* genes, and open up a number of new avenues for investigation, including hatch rates with additional controls to confirm the recapitulation of CI, assays to test self-rescue, qPCR experiments to compare wild type gene and transgene expression, and immunohistochemistry to test whether the CI seen here accurately reflects the cytological defects seen in *Wolbachia*-induced CI. These experiments and associated preliminary data will be further discussed in chapter VII. If confirmed as mediators of CI, the discovery of these genes will begin to answer the decades-long question of how bacteria are able to manipulate the reproduction of their hosts in such a profound manner. Additionally, they will allow further investigation into the molecular pathways underpinning this manipulation.

These genes also have a practical application in that they could be used for the creation of transgenic insects, replacing or complementing the pest control strategies currently utilizing *Wolbachia* infections (Brelsfoard and Dobson, 2009). One particularly promising approach is the use of variants of the sterile insect technique (Knipling, 1959), in which male pest insects are made infertile, either through irradiation (Dyck et al., 2005), infection with CI-causing strains of *Wolbachia* (Bourtzis et al., 2014, O'Connor et al., 2012), or introduction of a dominant lethal transgene (Leftwich et al., 2014). The result in each case is decreased fertility or death of offspring and a drastic drop in pest populations when significant numbers of modified males are released into the wild. These techniques have shown much promise, and successfully eradicated screwworms from North America (Dyck et al., 2005) and the melon fly from much of Japan (Dhillon et al.,

2005). However, irradiation can reduce fitness of the released males (Collins et al., 2008, Norikuni et al., 2008), and some of the most important mosquito disease vectors have been unable to be stably infected with *Wolbachia* (Hughes et al., 2011), or have undesirable effects on vector competency (Zele et al., 2014, Dodson et al., 2014). A version of this technique using CI transgenes would be unlikely to suffer from such issues. Transgenic insects could be created that express one or more CI effectors, and males would be released into a population. Since the wild females lack the CI rescue factor, all matings with these transgenic males would have dramatic reductions in offspring viability, crashing or even eliminating the pest insects from an area. Clearly much more remains to be done, but these experiments may be just the beginning of an exciting new chapter in understanding *Wolbachia* biology and host-microbe interactions, and in controlling a wide range of arthropod pests.

CHAPTER VI. MICROEVOLUTION AND INFECTION DYNAMICS OF *WOLBACHIA PIPIENTIS* AFTER HOST SWITCHING IN CELL CULTURE**

Abstract

After millions of years of co-evolution, vertically transmitted endosymbiotic bacteria are generally well adapted for life within a specific host. These endosymbionts have well-controlled population levels within their hosts, such that the infection is reliably transmitted to the next generation, but with minimal adverse effects on host fitness. Endosymbionts that switch their hosts, such as *Wolbachia pipientis*, typically have initial barriers to horizontal transmission, but have repeatedly overcome these barriers to infect new species throughout evolutionary history. Little is known about the genetic changes that occur during host switching, or the bacterial mechanisms that control endosymbiont population levels within a host. We show here through whole genome sequencing that host switching of the *Wolbachia* strain *w*Mel from *Drosophila melanogaster* fruit flies to mosquito cell culture resulted in a single nucleotide change in the *w*Mel gene encoding its RNase P protein component, but this mutation was not fixed in the population. The mutation may have functional consequences for processing of the 4.5S RNA, which controls the trafficking of secreted proteins. Additionally, in a host-switching experiment between fruit fly and mosquito cell cultures, both the previous host and the new host influenced infection progression through unknown mechanisms. These

---

** William Martin, Minhee Jo, Nicholas Reiter, and Seth R. Bordenstein contributed to the authorship of this chapter.

experiments provide new insight into the processes involved in host adaptation in endosymbionts and establish new avenues for investigation of these phenomena.

Introduction

Endosymbiotic bacteria undergo extensive co-evolution with their eukaryotic hosts, often resulting in the loss of many genes necessary for a free-living life style or for symbiosis with alternative hosts (Klasson and Andersson, 2004). This is the case with an endosymbiont of aphids, *Buchnera aphidicola*, a bacterium with strictly vertical transmission whose minimalist genome has been relatively static for millions of years after an initial rapid reduction in genome size and loss of genes that are non-essential for its intracellular life style (van Ham et al., 2003). Other intracellular species, such as *Wolbachia pipientis*, retain the ability to be transmitted horizontally, switching from one host species to another (Vavre et al., 1999), and have a more dynamic genome reflective of this changing environment and exposure to novel gene pools (Ishmael et al., 2009, Metcalf and Bordenstein, 2012).

Although the mechanisms for *Wolbachia* host switching in nature are only partially understood (Le Clec'h et al., 2013, Heath et al., 1999), horizontal transmission is common over evolutionary timescales (Zhou et al., 1998, Baldo et al., 2006) and many *Wolbachia* strains have been experimentally transferred between hosts in the laboratory (Rigaud et al., 2001, Riegler et al., 2004, Xi et al., 2005). The *Wolbachia* strain *w*MelPop, for example, has been successfully transferred from *Drosophila melanogaster*, where it was first discovered (Min and Benzer, 1997), to mosquito cell culture and then to *Aedes aegypti* mosquitoes (McMeniman et al., 2009). Transmission from *D.*

melanogaster to *A. aegypti* could not be accomplished directly, requiring years of cell culture adaptation (thereafter referred to as "cell line adapted" or *w*MelPop-CLA) prior to injection into mosquitos. Interestingly, when *w*MelPop-CLA was microinjected back into *D. melanogaster* embryos, bacterial titers and virulence were drastically reduced (McMeniman et al., 2008). These data suggest that an evolutionary change occurred following transfer to cell culture that made *w*Mel more fit for life in mosquito cells and less fit for infection of its original host. *w*MelPop and *w*MelPop-CLA have been recently sequenced, identifying five genetic changes associated with cell culture adaptation to a novel host (Woolfit et al., 2013), though the functional consequences of these mutations remain unclear.

Bacterial populations typically remain relatively stable in endosymbionts in order to maintain their symbiotic relationship with their host. For instance, if titers become too low, *Wolbachia* may be unable to successfully manipulate its hosts' reproduction or may fail to be transmitted to the next generation (Breeuwer and Werren, 1993, Dyer et al., 2005, Dutton and Sinkins, 2004, Jaenike, 2009). On the other hand, if population levels are too high, the infection may become virulent leading to reduced host fitness and premature death (Min and Benzer, 1997). Thus, a balance must be maintained in bacterial symbiont populations to preserve evolutionary fitness of the host-symbiont system. Although it is thought that the host immune system may play a role in regulating endosymbiont titers (Reynolds and Rolff, 2008), little is known about the bacterial genes involved in this process. To evaluate endosymbiont genes associated with bacterial titer and to examine genomic changes following cell culture adaptation to a novel host, we sequenced *Wolbachia* both in its native host and after cell culture adaptation, and

followed its infection dynamics over time after switching hosts. We identified a unique single nucleotide polymorphism (SNP) in *w*Mel *Wolbachia* appearing after adaptation to mosquito cell culture, resulting in an amino acid change in the RNase P protein component. We show that both donor and recipient hosts of transferred *Wolbachia* influence infection progression. These results shed light on the microevolutionary changes after an endosymbiont undergoes host switching and may reveal possible bacterial regulatory mechanisms controlling endosymbiont population dynamics.

Materials & Methods

Fly and cell lines

D. melanogaster strain yw, which natively harbors *w*Mel *Wolbachia*, uninfected RML12 *A. aegypti* cell lines, and RML12 lines infected with *w*Mel from yw flies were provided by Scott O'Neill (Monash University, Australia). Uninfected S2R+ *D. melanogaster* cells were provided by Irene Newton (Indiana University, USA). Flies were reared on conventional fruit fly food while cell cultures were maintained through weekly passages in Schneider's media with 10% FBS and penicillin/streptomycin for approximately seven years prior to these experiments.

Whole genome sequencing

*w*Mel was purified from four T75 flasks of RML12-*w*Mel cells as previously described (Gamston and Rasgon, 2007), while ten whole yw *D. melanogaster* females were frozen in liquid nitrogen and DNA was extracted from all samples using a PureGene

DNA purification kit (Qiagen, Venlo, Netherlands). One μg of DNA from each sample was sonicated, adapter-ligated, and sequenced using 100bp paired-ends on an Illumina Hi-Seq (Vanderbilt Sequencing Core, Nashville, TN). Using CLC Genomics Workbench version 6.0.4 (CLC Inc., Aarhus, Denmark), sequencing reads were trimmed and mapped to the *w*Mel reference genome (Wu et al., 2004) or assembled *de novo*. Approximately 26 million reads were generated for *w*Mel-yw and 17 million reads for *w*Mel-RML12. 6.3% of *w*Mel-yw reads mapped to the reference *w*Mel genome for an average coverage of 117, while 70.8% of *w*Mel-RML12 reads were mapped for an average coverage of 864. SNP and DIP variations were detected with CLC probabilistic and quality-based variation algorithms, and large structural variation such as rearrangements and large insertions and deletions were examined with the structural variant detection tool. To detect any novel genes, de novo contigs were mapped to the reference genome with a permissive length fraction of 0.25 and any overhangs that did not match the reference were examined manually. All variants present in 15% or more of the raw reads were confirmed with PCR and Sanger sequencing. Assembly of Sanger sequencing results, BLAST searches, and multiple sequence alignments with the MUSCLE algorithm (Edgar, 2004) were performed using Geneious Pro 5.5.6.

wMel culture experiments

*w*Mel was transferred between hosts using a modified shell-vial technique (Rasgon et al., 2006). In culture experiment 1, *w*Mel was purified from RML12 cells or yw *D. melanogaster* embryos as previously described (Gamston and Rasgon, 2007) and inoculated onto confluent monolayers of uninfected RML12 or S2R+ cells in shell vials.

Both hosts for a given purification received an identical dose of *w*Mel. Vials were centrifuged for one hour, allowed to incubate overnight at 26 °C, then transferred to 6-well plates. After 2 days, cells were transferred to T25 culture flasks and then passaged 1:10 every week thereafter. 200 mL samples were collected prior to each passage for detecting bacterial titer. *w*Mel titer was examined with qPCR by comparing copy number of the *Wolbachia* GroEL gene to the *D. melanogaster* Actin5C gene or *A. aegypti* ribosomal protein S7. All primers are listed in Table VI-1. In culture experiment 2, *w*Mel was purified from cultured cells that received host-matched *w*Mel in experiment 1. 8 x 10^6 *w*Mel cells were added to 2.35 x 10^6 host cells that had been split and allowed to rest overnight and cultured as described above. Cells were counted with a hemocytometer and counting of *w*Mel was facilitated with a BacLight Live/Dead Stain (Invitrogen). Prevalence of the WD_0200 mutation was monitored by PCR and sequencing, followed by population analysis using Mutation Surveyor (Softgenetics, State College, PA, USA).

Table VI-1. Primers used in this study. S7 primers derived from (Xi et al., 2008b).

Target	Forward Primer (5'-3')	Reverse Primer (5'-3')
185816 (*w*Mel)	TGAAGGAGAATTACTGACTTCTGGT	AAAACCGGAAAAGCAGCAAAAAGA
1006082 (*w*Mel)	TGGGCAGGTTCTGCTTTTAGTTTTA	TCCTGCAGTTTATTTCTCATGTGCT
1020476 (*w*Mel)	GGGATGACATAAATAAGCGCTGAAA	ACATGATCCGCAAGGTTCAAAAAT
1094459 (*w*Mel)	TGCTTTATAAGCGCATTTTAGCTTG	AAGAAAAAGGGTACAGCGTGTCTAA
1097797 (*w*Mel)	TTTGCCTGCAAGCGAGGAATTATTA	AGAGTGAACATCACAAACTTTTGCTT
1103469 (*w*Mel)	ATATACTGAAAGAACGCGGCGAATA	AGCTTCAACAACCTTGCTTATACGA
1161851 (*w*Mel)	GCTGAATCTGCAAATTTTCCGTGTA	AACTGCCATCACAACAATTGACAT
1163171 (*w*Mel)	TGGCTTTATAACACCAACCGAATCA	TTGTTGAAATCATGCGCAATACCA
1177854 (*w*Mel)	TTCCTCCATTTCATCAACATTTGCC	AGCGCAATAAAACACTGCAAACAT
GroEL (*w*Mel)	CAACCTTTACTTCCTATTCTTG	CTAAAGTGCTTAATGCTTCACCTTC
Actin5c (*D. melanogaster*)	ATGTGTGACGAAGAAGTTGCT	GTCCCGTTGGTCACGATACC
Ribosomal protein S7 (*A. aegypti*)	GGGACAAATCGGCCAGGCTATC	TCGTGGACGCTTCTGCTTGTTG

Results

Minimal genomic changes following introduction to cell culture

Given that transfer of *Wolbachia* from *D. melanogaster* flies to *A. aegypti* cell culture represents a change in both host species and growth conditions, and the fact that this transition is associated with reduced ability to reinfect the original host after only a few years in culture, we hypothesized that the switch would be associated with genetic adaptation. To examine this possibility, we performed whole genome sequencing on *Wolbachia* from *A. aegypti* cell culture (*w*Mel-RML12) and on *Wolbachia* in the *D. melanogaster* strain from which the cell culture line was derived (*w*Mel-yw). Genomes were mapped to the reference *w*Mel genome (Wu et al., 2004) and examined for sequence variation between the two strains and the reference. Sequencing depth was robust, with an average 117-fold coverage for *w*Mel-yw and 864-fold coverage for *w*Mel-RML12. Neither major structural variations such as insertions/deletions or inversions, nor novel gene insertions were detected, but a total of seven deletion/insertion polymorphisms (DIPs) and two single nucleotide polymorphisms (SNPs) compared to the reference genome were noted and confirmed with PCR and Sanger sequencing (Table VI-2). Of these variations, only the T->A SNP at reference position 185,816 was unique to *w*Mel-RML12, while the remaining DIPs and SNP were shared between *w*Mel-RML12 and *w*Mel-yw. Additionally, when compared to a resequencing compilation of *w*Mel from 179 *D. melanogaster* strains (Richardson et al., 2012), this SNP was not present in any laboratory *w*Mel strain, while remaining SNP and DIPs were present in at least 6.7% and up to 95.5% of the sequenced strains (Table VI-2). The SNP at reference position 185,816 was not present in cells that had been frozen after four years of culture,

indicating that the mutation likely appeared after four but before seven years of cell culture (data not shown).

wMel-RML12 mutation is in a conserved region of RNase P protein

The T->A SNP at reference position 185,816 in *w*Mel-RML12 results in an amino acid change from leucine to phenylalanine at position 41 in the protein WD_0200, which is annotated as a hypothetical protein. A protein BLAST search using the amino acid sequence of WD_0200 and multiple sequence alignment of matches revealed that the protein is highly conserved across sequenced *Wolbachia* strains including those from other *Wolbachia* supergroups (Figure VI-1). Leu41 is conserved in all of these strains and there are no mutations in the gene in any of the 179 resequenced *w*Mel genomes (Richardson et al., 2012). Comparison with annotation in other *Wolbachia* strains and other bacterial species suggests that WD_0200 is the protein component of RNase P, a ribozyme involved in processing and maturation of tRNA and a number of small non-coding RNAs (Marvin and Engelke, 2009). Interestingly, a SNP in the same gene was noted as one of the five genomic changes associated with cell line adaptation in the related *Wolbachia* strain *w*MelPop (Figure VI-1) (Woolfit et al., 2013), an extremely unlikely occurrence by chance alone.

Table VI-2. Genetic variation before and after *w*Mel host-switching. Genetic changes in *w*Mel from *Aedes* cell culture (*w*Mel-RML12) compared to native *Drosophila* infection (*w*Mel-yw), the *w*Mel reference strain, and the DGRP/DPGP *w*Mel resequencing project (Richardson et al., 2012). All variants were present in more than 85% of *w*Mel-RML12 sequencing reads. The only variant that is different between *w*Mel-RML12 and *w*Mel-yw is shaded.

Reference Position	Variant Type	*w*Mel-RML12	*w*Mel- yw	*w*Mel- ref	% Presence in DGRP/ DPGP	Gene(s) Affected	Putative Function	Amino Acid Change
185816	SNP	A	T	T	0.0%	WD_0200	RNase P protein component	Leu41Phe
1006082	DIP	T	T	-	56.4%	WD1043, WD1044	Conserved hypothetical proteins	Cys317fs, Val1fs
1020476	DIP	T	T	-	58.7%		Intergenic region	
1094459	DIP	T	T	-	70.4%		Intergenic region	
1097797	SNP	A	A	T	95.5%		Intergenic region	
1103469	DIP	-	-	T	14.0%	WD1155	Conserved hypothetical protein	Phe95fs
1161851	DIP	T	T	-	6.7%		Intergenic region	
1163171	DIP	-	-	C	91.1%	WD1215	Sensor histidine kinase	Ser351fs
1177854	DIP	-	-	C	32.4%	WD1231	Protoheme IX biogenesis protein	Tyr123fs

Figure VI-1. Conservation of WD_0200, a putative RNase P protein component, between sequenced strains of *Wolbachia*. Highly conserved residues are indicated with black or grey boxes. Sequence logo shows amino acids present at each locus with the size of the amino acid symbol reflecting its frequency at that position in all sequences. Identity plots the conservation of each residue, with tall green bars indicating absolute conservation and yellow bars indicating less conservation. Mutations found in *w*MelPop-CLA (Woolfit et al., 2013) and *w*Mel-RML12 (this study) are emphasized with red boxes.

Infection dynamics during wMel host switching

To evaluate whether this mutation was associated with any changes in *w*Mel infection dynamics, we performed a host-switching experiment in which RML12 *A. aegypti* or S2R+ *D. melanogaster* cells were infected with either the mosquito-adapted *w*Mel-RML12 or the fruit fly-adapted *w*Mel-yw. Interestingly, titer dramatically increased in S2R+ cells infected with *w*Mel-RML12 compared to other lines (Figure VI-2A), before decreasing back to a level similar to other cultures after four months. Additionally, when *Wolbachia* were moved from yw flies to either species in cell culture, bacterial titer lagged significantly before reaching a similar titer to other lines after four months. These variations were statistically dependent on both the donor and recipient hosts. To determine whether this phenotype could be due to variation in the numbers of

host cells or *w*Mel cells in the initial infection or to an artifact of moving from infection in whole flies to cells in culture, we repeated the experiment using identical numbers of cells across all cultures and with *w*Mel purified from S2R+ cell cultures rather than yw flies. Again, *w*Mel-RML12 in *D. melanogaster* host cells reached a markedly higher titer than the other three cell lines and variation in titer was statistically dependent on both donor and recipient hosts (Figure VI-2B). However, this variation in titer was independent of the mutation detected in WD_0200, as the prevalence of the SNP dropped from 86% of Illumina reads, to an average of 52% of the population in experiment #1 and <5% of the population in experiment #2.

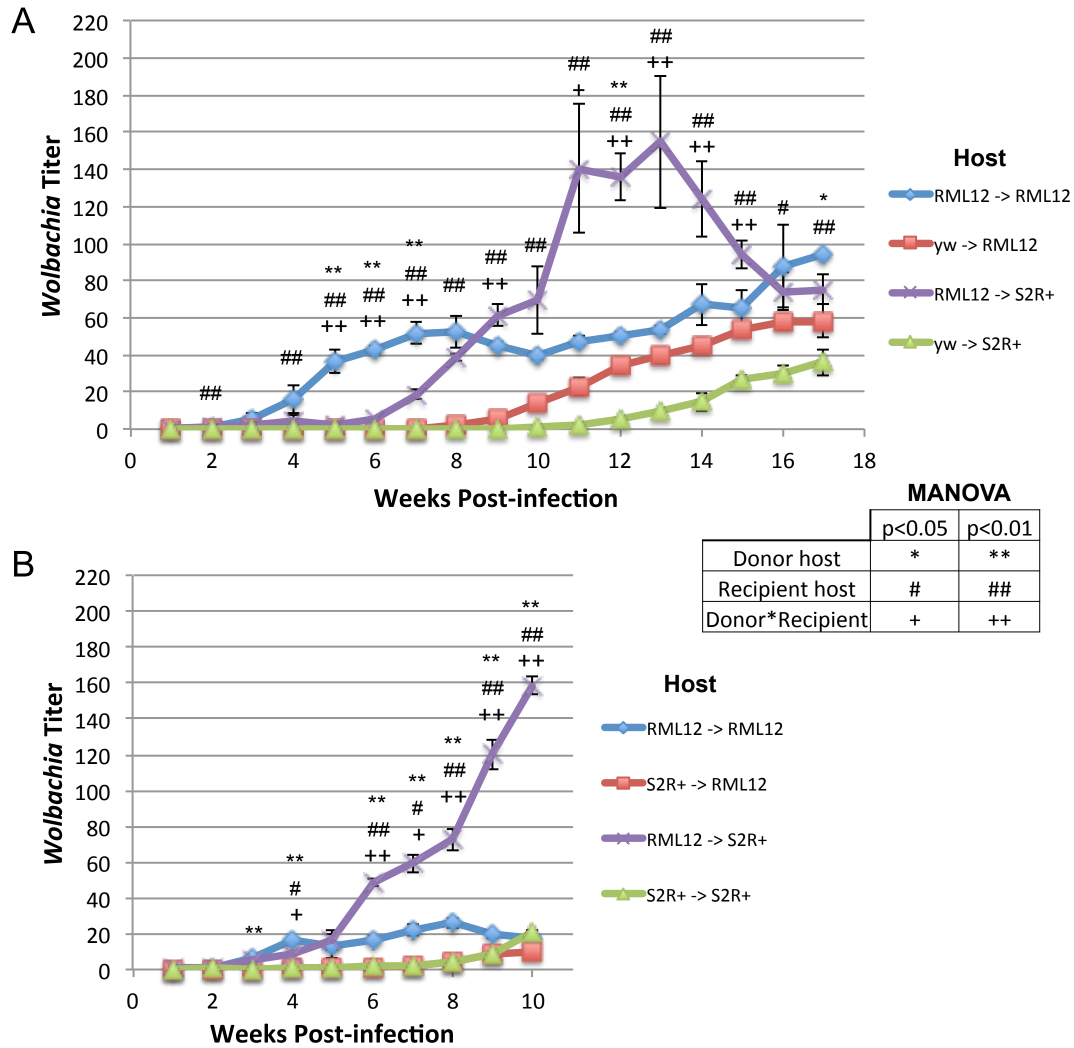


Figure VI-2. *w*Mel infection dynamics after switching or maintaining eukaryotic hosts. Titer = copy number of *wMel* GroEL gene per copy of *D. melanogaster* Actin5C or *A. aegypti* S7 gene as determined via qPCR. **(A)** Experiment #1; N = 3 for all lines. **(B)** Host-switching experiment #2; N=4 for all lines. Error bars are +/- standard error of the mean.

Discussion

Minimal genetic change associated with host switching

Given the alteration in phenotype (McMeniman et al., 2008) and relatively rapid evolution (Woolfit et al., 2013) seen in *w*MelPop cell line adaptation, we expected to discover a number of genetic mutations in *w*Mel after extended adaptation to culture in

mosquito cells. On the contrary, only a single SNP was detected, in the putative RNase P protein component WD_0200. It is possible that the increased virulence and growth rate seen in *w*MelPop (Min and Benzer, 1997) favored more rapid evolution and stronger selection than the less proliferative *w*Mel. The single mutation in *w*Mel was not fixed in the *Wolbachia* population, as its prevalence dropped from 86% of reads in the Illumina sequencing run, to less than 5% of the population in the second host-switching cell culture experiment 18 months later. There are at least two possible explanations for this result. One is that the mutation only occurred as a product of genetic drift, without a selective benefit for the bacterium. The second is that the mutation did have a selective benefit, but this advantage was mitigated by the highly regimented culture conditions of the host-switching experiments, resulting in its gradual loss over a period of months. The latter possibility is tempting given the unlikely occurrence of a mutation in the same gene after cell line adaptation in *w*MelPop (Woolfit et al., 2013). Interestingly, RNase P is required for cleavage of not only pre-tRNA, but also pre-4.5S RNA, or signal recognition particle RNA (Esakova and Krasilnikov, 2010, Guerrier-Takada and Altman, 1984, Peck-Miller and Altman, 1991), and substrate specificity is determined in large part by the protein component of the ribozyme (Marvin and Engelke, 2009, Peck-Miller and Altman, 1991, Liu and Altman, 1994). The 4.5S RNA is a universally conserved RNA component of a ribonucleoprotein that regulates trafficking of integral membrane proteins targeted for secretion to the plasma membrane or periplasm (Saraogi and Shan, 2014, Akopian et al., 2013). Thus, one potential mechanism of cell culture adaptation would be an altered repertoire of surface proteins due to changes in 4.5S RNA processing. Determining whether or not this is the case will require functional tests of the RNA processing

capabilities of wild type and mutant WD_0200 in conjunction with the RNase P RNA component and these experiments are ongoing.

Host-dependence of cell culture infection progression

Since the variation in *Wolbachia* titer in *A. aegypti* and *D. melanogaster* cells (Fig. 2) depended in part on the donor host but was independent of the *Wolbachia* WD_0200 gene mutation, a non-genetic mechanism for these variations and perhaps the phenomenon of cell line adaptation in general must be postulated. One possible explanation is that of epigenetics. Although traditionally viewed as a eukaryotic phenomenon, epigenetic modifications can also have profound effects on bacterial phenotypes and can be stably transmitted between generations (Casadesus and D'Ari, 2002, Casadesus and Low, 2013, Casadesus and Low, 2006). These mechanisms of memory include autocatalytic protein loops, carbohydrate modifications of the cell wall, and perhaps most notably, DNA methylation. Indeed, methylation can drastically change gene expression patterns (Fang et al., 2012), alter virulence phenotypes (Heithoff et al., 1999), and control progression through the cell cycle (Reisenauer et al., 1999). Thus, determination of the transcriptome and methylome of wild type and cell line adapted *Wolbachia* may further elucidate the mechanisms associated with adaptation to novel hosts.

CHAPTER VII. CONCLUSIONS AND FUTURE DIRECTIONS

The research described in this thesis has made substantial progress in understanding several aspects of *Wolbachia* biology, including the role of phage WO in *Wolbachia*-host interactions, evolution and horizontal transfer of WO genes, application of horizontally transferred WO gene homologs, and mechanisms of *Wolbachia* adaptation to novel hosts. As with most scientific undertakings however, the results of these studies raise even more questions and reveal a number of fascinating avenues for future investigations. Some of these questions and future directions are discussed below.

Phage WO mechanisms of lysis

One interesting biological question that remains unanswered is how WO causes lysis of *Wolbachia* during the lytic phase of its life cycle. This question is of substantial interest, because although the lytic mechanisms of phages in free-living bacteria are fairly well understood (Young, 2013), there has been little study of these processes in phages infecting obligate intracellular microbes. Indeed, it was the investigation of WO lysozyme as a tool for phage-induced *Wolbachia* lysis that led us to investigate its evolution and horizontal transfer across the tree of life, discussed in Chapter III. Host lysis presents a special challenge for phage WO, as it must not only break through the inner and outer cell membrane, but presumably at least a rudimentary cell wall as well as multiple Golgi-derived eukaryotic membranes that commonly encapsulate the bacteria (Cho et al., 2011).

WO lysis candidates

Bacteriophages encode numerous products that inhibit or manipulate their hosts, but generally only a few proteins are directly required for host lysis (Young et al., 2000, Liu et al., 2004, Shibayama and Dabbs, 2011). At least six WO haplotypes produce active virions (Bordenstein et al., 2006, Fujii et al., 2004, Gavotte et al., 2007, Kent and Bordenstein, 2010, Sanogo and Dobson, 2006), with lethal effect on *Wolbachia* through unknown mechanisms. Transmission electron micrographs demonstrate typical cellular defects associated with phage-induced lysis including degraded DNA, collapsed outer membranes, and phage particle exit (Bordenstein et al., 2006). We have identified four gene products encoded by WO that could be mediators of *Wolbachia* lysis: patatin, holin, lysozyme, and spanins.

Patatin is presumably of crucial importance in the WO life cycle as it is present in all WO prophages with tail modules, which are necessary for active virion production (Kent et al., 2011a). Patatin is in the phospholipase A2 (PLA2) enzyme family (Osborne and Campbell, 1896, Shewry, 2003) and has activity against several types of lipids (Scherer et al., 2010). Interestingly, animal PLA2 is able to hydrolyze bacterial phospholipids (Nevalainen et al., 2008), suggesting that WO patatin could participate in *Wolbachia* lysis by destroying bacterial membranes. However, in some *Pseudomonas* and *Streptococcus* species, PLA2 is a virulence factor attacking eukaryotic membranes (Sitkiewicz et al., 2007). Patatins in phage genomes are a rarity and the WO patatin is too divergent from characterized enzymes to definitively predict what specificity it may have without a functional approach.

Holins are part of an extremely diverse group of pore-forming proteins commonly produced by dsDNA phages (Wang et al., 2000). Holins accumulate in bacterial membranes and oligomerize to form large pores causing disruption of membrane potential and cell death (Savva et al., 2008, White et al., 2011). A putative holin is present in several, but not all WO phages, although annotation of holins can be exceedingly difficult due to the large diversity and minimal sequence similarity between individual holins. While holins have antibacterial potential alone (Shi et al., 2012), complete bacterial lysis leading to bacteriophage release is usually accomplished with a lysozyme, which passes through the holin pore to digest cell wall peptidoglycan (Young et al., 2000).

At least three WO haplotypes have a canonical lysozyme, a GH25 muramidase that is found in many phages of free-living bacteria where it digests the peptidoglycan cell wall as the key player in their lysis cassette. Thus, the lysozyme is an obvious lysis candidate, however, it is not present in all WO haplotypes known to produce active virions, suggesting it may not be crucial to *Wolbachia* lysis. Additionally, although the production of cell wall precursors are required for *Wolbachia* viability, traditional cross-linked peptidoglycan has not been detected in *Wolbachia*, and if it does contain a cell wall it is likely thin and atypical in structure.

The final candidate proteins are spanins, essential components of the lysis cassette in phages infecting Gram-negative bacteria required, where they disrupt the outer bacterial membrane (Berry et al., 2012, Summer et al., 2007). Although spanins have not been annotated in phage WO, we identified two proteins with diagnostic characteristics of previously annotated spanins (Summer et al., 2007). These characteristics include a

putative outer membrane lipoprotein, designated by the presence of a signal peptide, in the first spanin (Sp1) and a coiled coil protein with a single amino-terminal transmembrane domain in the second putative spanin (Sp2). C-terminal binding of the lipoprotein with the coiled-coil domain forms the spanin complex that "spans" the distance between the inner membrane and outer membrane. The activity of coiled-coil transitions in membrane fusion events is a promising hypothesis for the membrane disruption potential of spanins. Like the patatin, the spanins are widespread in WO haplotypes.

WO lysozyme is likely not solely responsible for Wolbachia lysis

As a first step to investigating WO lysis mechanisms, we cloned the lysozyme from WORiA infecting *w*Ri in *Drosophila simulans* and expressed the recombinant protein in either *E. coli* for purification and *in vitro* studies or in *D. melanogaster* for *in vivo* testing of anti-*Wolbachia* activity. WORiA lysozyme was cloned into a pET20b-6xHis vector, expressed, and isolated in high purity with a nickel affinity column (Figure VII-1A). When purified lysozyme was added in concentrations up to 100 mg/mL to *w*Mel *Wolbachia* isolated from host cells in culture, no increase in cell death was apparent over a negative control of bovine serum albumin (Figure VII-1B). These data suggest that lysozyme is likely not able to lyse *Wolbachia* alone, although a number of other possibilities could account for this result, including inhibition of lysozyme activity by culture media, inability to lyse from outside the *Wolbachia* cell, protection of *Wolbachia* by host Golgi-derived membranes not removed in the isolation procedure, or necessity of additional lysis components.

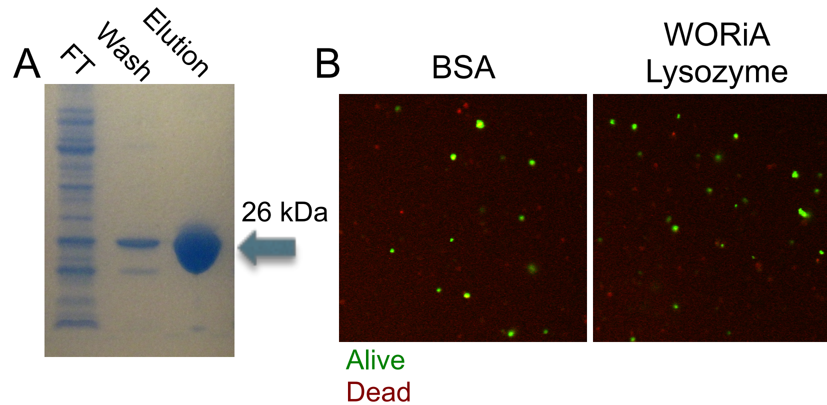


Figure VII-1. Testing WORiA lysozyme activity in vitro. (A) Coomassie blue stained PAGE gel showing flow through (FT), wash, and elution fractions from a purification of the 26 kDa WORiA lysozyme. (B) *Wolbachia* visualized at 40x using live/dead stain after treatment with 100 μg/mL bovine serum albumin (BSA) or WORiA lysozyme. Cells with intact membranes (live) are green, while cells with disrupted membranes (dead) are red.

For *in vivo* studies, WORiA lysozyme was cloned into a pUAST vector using the GAL4-UAS system for inducible expression and injected into *D. melanogaster* to create transgenic fly lines via P-element transposon recombination. The WORiA minor capsid protein C gene was also cloned and injected in an identical fashion to create a control line expressing a protein not thought to be involved in *Wolbachia* lysis. These transgenic flies were crossed with *w*Mel-infected geneswitch-Gal4-actin5c flies (Shen et al., 2009) to drive ubiquitous expression of the transgenes (confirmed with RT-qPCR, data not shown), inducible by addition of mifepristone to fly media and *Wolbachia* titer was monitored with qPCR. Similarly to the *in vitro* assays, expression of lysozyme had no effect on *Wolbachia* viability (Figure VII-2), further supporting that the gene cannot kill *Wolbachia* from outside the bacterium when acting alone. Again, a number of alternative explanations could explain this result, such as misfolding of the recombinant protein, poor expression in gonads, or trafficking to inappropriate cellular compartments.

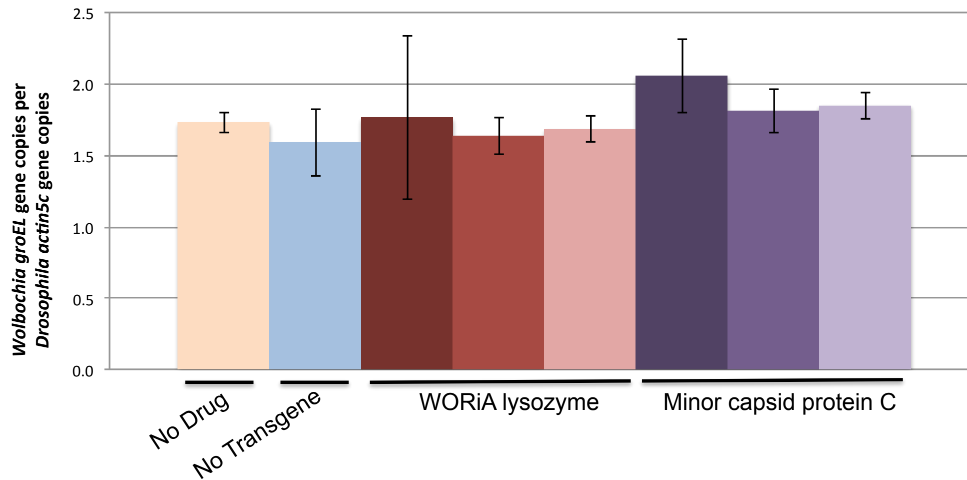


Figure VII-2. Testing WORiA lysozyme effect on *Wolbachia* titer *in vivo*. Titer determined by qPCR of *Wolbachia groEL* gene compared to *D. melanogaster actin5c* gene. Ten flies from each of three independent transgenic lines for both lysozyme and minor capsid protein C were assayed along with ten flies from wild type crosses with no transgene or uninduced transgene receiving no drug. No significant differences across any treatment group by one-way ANOVA.

Proposed mechanism of WO lysis

The preliminary data shown above suggests that lysozyme may not be able to lyse *Wolbachia* alone, indicating the other candidates may be of greater importance. Interestingly, it was recently demonstrated that when peptidoglycan crosslinks are removed, recapitulating the likely structure of the *Wolbachia* cell wall (Vollmer et al., 2013, Henrichfreise et al., 2009), spanins can cause cell rupture independently of any other lysis cassette components (Manoj Rajaure and Ry Young, Texas A&M University, unpublished data). In combination with the patatin digesting host-derived membranes, these proteins might be able to cause complete lysis of *Wolbachia* and host vacuoles (Figure VII-3) and would explain while they are so conserved across WO haplotypes. Investigation of this hypothesis will require recombinant expression of these proteins

alone and in concert, preferably in an inducible system using a *Wolbachia*-infected insect cell line.

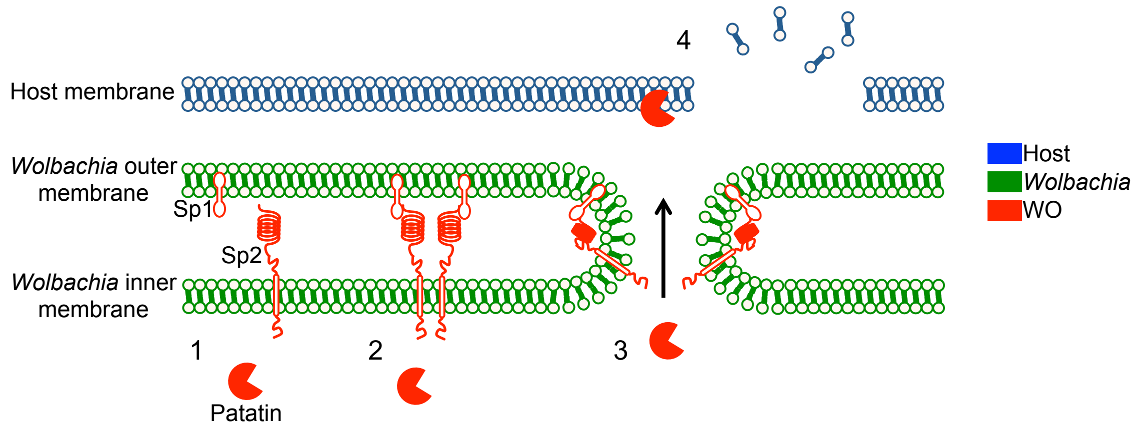


Figure VII-3. Proposed mechanism of *Wolbachia* lysis by WO. (1) Spanin protein 1 (Sp1) localizes to the outer *Wolbachia* membrane, while spanin protein 2 (Sp2) is in the inner *Wolbachia* membrane and patatin is intracellular. (2) Sp1 and sp2 associate and dimerize (Berry et al., 2013). (3) The coiled coil domain of Sp2 undergoes a conformational change, bringing the inner and outer *Wolbachia* membranes together in the absence of crosslinked peptidoglycan. This forms a pore in *Wolbachia* allowing patatin to exit the bacterial cell. (4) Patatin cleaves host-derived membranes, fully lysing the *Wolbachia*-host membrane complex and releasing phage progeny.

Antibacterial activity of WO lysozyme homologs

A sharp decline in antibiotic drug development has led to only a handful of new antibacterial medications reaching the market in the past decade. Meanwhile, antibiotic overuse has selected for rapid evolution of multi-drug resistant organisms, making the need for alternative therapeutics all the more urgent (Boucher et al., 2009). Unfortunately, most drugs under development are only incremental improvements on old antibacterial compounds with similar specificities, related modes of action, and a high risk for cross-resistance (Theuretzbacher, 2012). For these reasons, development of drugs

with novel mechanisms is imperative. The family of horizontally transferred lysozymes described in chapter III is one potential source of new antibacterial therapies.

The lysozyme in the archaeon *Aciduliprofundum boonei* is of particular interest as it has demonstrated antibacterial activity (Figure III-11, Figure III-13). Archaea have been largely ignored in the context of human health applications, likely because no pathogenic archaea have been identified (Eckburg et al., 2003). However, given that archaea coexist with bacterial species in the environment (Oren, 2002, Kato and Watanabe, 2010, Orcutt et al., 2011a) and can compete for similar resources, there may be significant unexploited potential for antibiotic compound discovery in this domain. While the GH25 muramidase domain alone was limited in activity to two families of Firmicutes, the full-length lysozyme was toxic to the *E. coli* it was expressed in suggesting a broader specificity. It is possible that individual domains or combinations of domains will have a spectrum of activity that could be useful as specific or semi-specific antibacterial peptides (Figure VII-4).

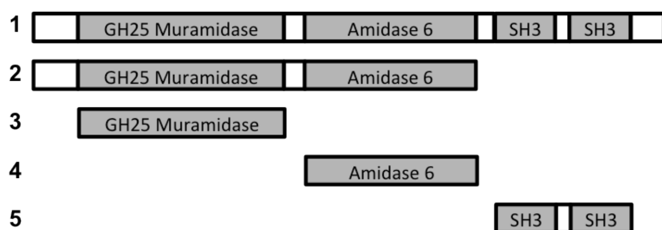


Figure VII-4. Proposed *A. boonei* lysozyme constructs. To evaluate the contribution of individual lysozyme domains to antibacterial activity, five constructs will be generated. Construct 1 and 3 have already been tested and exhibit activity against the *E. coli* in which the construct was expressed or against two families of Firmicutes, respectively.

Other horizontally transferred lysozymes may also be of interest as antibacterial agents. However, preliminary efforts to purify lysozymes from the plant *Selaginella moellendorffii*, and the aphid *Acyrthosiphon pisum*, have been unsuccessful due to insolubility of the recombinant proteins (Figure VII-5). Both lysozymes were successfully expressed with 6x-His tags, but were unable to be solubilized from the cell pellet, even with addition of detergents. Solubility was also a problem when the *S. moellendorffii* lysozyme was expressed in Sf9 insect cells (data not shown). Future purification experiments will require alternative tags that may improve solubility, such as mannose-binding protein.

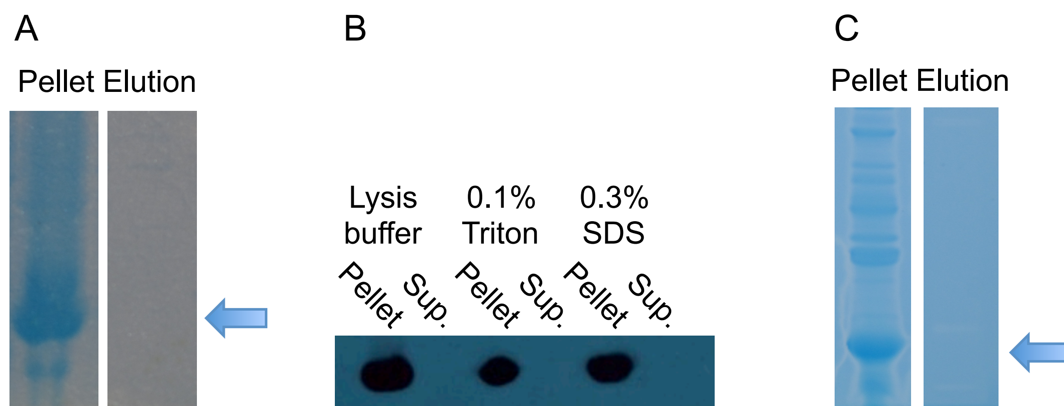


Figure VII-5. Insolubility of 6x-His tagged horizontally transferred lysozymes. (A) Expression of *S. moellendorffii* lysozyme. Coomassie blue-stained PAGE gel of cell pellet and elution fraction after affinity chromatography are shown with expected lysozyme band indicated with an arrow. (B) Western blot of cell pellets and supernatants (sup.) of *S. moellendorffii*-expressing cells using an anti-his antibody, after treatment with the indicated detergents. SDS = sodium dodecyl sulfate. (C) Coomassie blue-stained PAGE gel showing expression and purification of *A. pisum* GH25 muramidase domain with expected protein band indicated.

Mechanism of cytoplasmic incompatibility

As the first experiments to show modification of insect hatch rates by *Wolbachia* transgenes, our preliminary data on effectors of cytoplasmic incompatibility (CI)

described in Chapter V offer an exciting opportunity to investigate this poorly understood phenomenon. To verify these results we will repeat all hatch rate assays that demonstrated statistically significant differences in early experiments, using replicates of 32-48 matings of *Drosophila melanogaster* per cross and appropriate controls (Table VII-1), as well as collecting embryos for cytology and testes for qPCR, all in a single experiment for each gene. These crosses will establish whether transgenes are able to induce CI and its relative effect compared to *Wolbachia*-induced CI, if *Wolbachia* can rescue transgene-induced CI, whether the transgenes can rescue *Wolbachia*-induced CI, and if the transgenes are able to effect self-rescue. Embryo cytology will enable determination of whether transgenes cause the same chromosomal and cell cycle defects as in *Wolbachia*-induced CI. Histology will be performed using anti-phosphorylated histone H3 antibody and propidium iodide as previously described (Tram and Sullivan, 2002) on wild type, transgene CI, and *Wolbachia* CI crosses set up at the same time as the hatch rate assays, using siblings of the flies used in those crosses. After flies have mated for 36 hours, males from *Wolbachia* CI and transgene CI crosses will be removed from hatch rate chambers and their testes will be dissected and preserved in RNAlater (Qiagen). RNA will be extracted and RT-qPCR will be performed to compare expression of the gene in *Wolbachia* to transgene expression, with the hypothesis that hatch rates and transgene expression will be inversely correlated.

Table VII-1. Proposed CI experiment crosses. Key: +: infected with *w*Mel *Wolbachia*; -: uninfected; UAS: upstream activation sequence promoter driving transgene expression.

Cross #	Female	Male	Cross type
1	(-)	(-)	Wild type
2	(-)	(+)	*Wolbachia* CI
3	(+)	(+)	*Wolbachia* rescue
4	(-)	(-) UAS	Transgene CI
5	(+)	(-) UAS	*Wolbachia* transgene rescue
6	(-) UAS	(+)	Transgene rescue
7	(-) UAS	(-)	Female transgene control
8	(-) UAS	(-) UAS	Transgene self-rescue

***Wolbachia* adaptation to new hosts**

As described in chapter VI, when *w*Mel *Wolbachia* is cultured in mosquito cells *in vitro*, a single nucleotide polymorphism appeared in WD0200, a putative RNase P protein component. Although this mutation did not reach fixation in the population, the presence of a mutation in the same gene in *w*MelPop suggests that it may have a functional consequence for *Wolbachia* biology. To determine whether or not this is the case, we began experiments to express and purify recombinant WD0200 for use in RNA cleavage assays. WD0200 was cloned into three expression vectors attaching either a 6x-His tag, a glutathione S-transferase (GST) tag, or mannose binding protein (MBP) tag. While both the 6x-His tagged version and the GST tagged version were poorly soluble, the MBP tagged version was at least 50% soluble (Figure VII-6) and this version was selected for further study. Experiments are in progress to test this recombinant protein's ability to cleave pre-tRNA and pre-4.5S RNA and if wild type WD0200 is able to do so, both the *w*Mel-RML12 and *w*MelPop mutant versions will be tested to see whether this enzymatic activity is altered.

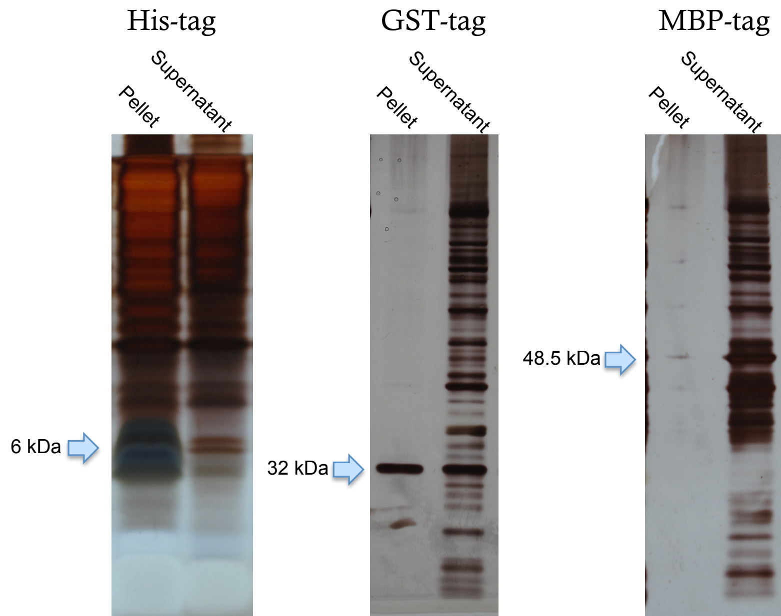


Figure VII-6. Solubility of recombinant wild type *w*Mel WD0200. Silver stained PAGE gels of pellet or supernatant from lysates of *E. coli* expressing either 6x-histadine tagged (His), glutathione S-transferase tagged (GST), or mannose-binding protein tagged (MBP) recombinant proteins. Arrows indicate expected size of each tagged protein.

Conclusions

Although there is still much work to be done, the research described here has built on existing studies to increase our understanding of the biology and multi-level symbioses of intracellular bacteria and their phages in general, and *Wolbachia* biology specifically. Additionally, the work promises several potential applications to human medicine, including new antibacterial therapeutics and novel methods of agricultural pest and disease vector control.

REFERENCES

ABASCAL, F., ZARDOYA, R. & POSADA, D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics,* 21**,** 2104-5.

AKOPIAN, D., SHEN, K., ZHANG, X. & SHAN, S. O. 2013. Signal recognition particle: an essential protein-targeting machine. *Annu Rev Biochem,* 82**,** 693-721.

AL-KHODOR, S., PRICE, C. T., KALIA, A. & ABU KWAIK, Y. 2010. Functional diversity of ankyrin repeats in microbial proteins. *Trends Microbiol,* 18**,** 132-9.

ALBERS, S. V. & MEYER, B. H. 2011. The archaeal cell envelope. *Nat Rev Microbiol,* 9**,** 414-26.

ANDAM, C. P. & GOGARTEN, J. P. 2011. Biased gene transfer in microbial evolution. *Nat Rev Microbiol,* 9**,** 543-55.

ANDAM, C. P. & GOGARTEN, J. P. 2013. Biased gene transfer contributes to maintaining the tree of life. *In:* GOPHNA, U. (ed.) *Lateral gene transfer in evolution.*

ANDERSSON, J. O. 2005. Lateral gene transfer in eukaryotes. *Cell Mol Life Sci,* 62**,** 1182-97.

ANDERSSON, J. O., SJOGREN, A. M., DAVIS, L. A., EMBLEY, T. M. & ROGER, A. J. 2003. Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. *Curr Biol,* 13**,** 94-104.

ANGIUOLI, S. V., MATALKA, M., GUSSMAN, A., GALENS, K., VANGALA, M., RILEY, D. R., ARZE, C., WHITE, J. R., WHITE, O. & FRICKE, W. F. 2011. CloVR: a virtual machine for automated and portable sequence analysis from the desktop using cloud computing. *BMC Bioinformatics,* 12**,** 356.

ATANASOVA, N. S., PIETILA, M. K. & OKSANEN, H. M. 2013. Diverse antimicrobial interactions of halophilic archaea and bacteria extend over geographical distances and cross the domain barrier. *Microbiologyopen,* 2**,** 811-825.

ATTAIECH, L., GRANADEL, C., CLAVERYS, J. P. & MARTIN, B. 2008. RadC, a misleading name? *J Bacteriol,* 190**,** 5729-32.

AUDOLY, G., VINCENTELLI, R., EDOUARD, S., GEORGIADES, K., MEDIANNIKOV, O., GIMENEZ, G., SOCOLOVSCHI, C., MEGE, J. L., CAMBILLAU, C. & RAOULT, D. 2011. Effect of rickettsial toxin VapC on its eukaryotic host. *PLoS One,* 6**,** e26528.

AUGUSTINOS, A. A., SANTOS-GARCIA, D., DIONYSSOPOULOU, E., MOREIRA, M., PAPAPANAGIOTOU, A., SCARVELAKIS, M., DOUDOUMIS, V., RAMOS, S., AGUIAR, A. F., BORGES, P. A., KHADEM, M., LATORRE, A., TSIAMIS, G. & BOURTZIS, K. 2011. Detection and characterization of *Wolbachia* infections in natural populations of aphids: is the hidden diversity fully unraveled? *PLoS One,* 6**,** e28695.

BALDO, L., DUNNING HOTOPP, J. C., JOLLEY, K. A., BORDENSTEIN, S. R., BIBER, S. A., CHOUDHURY, R. R., HAYASHI, C., MAIDEN, M. C., TETTELIN, H. & WERREN, J. H. 2006. Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Appl Environ Microbiol,* 72**,** 7098-110.

BALTRUS, D. A. 2013. Exploring the costs of horizontal gene transfer. *Trends Ecol Evol,* 28**,** 489-95.

BANDI, C., TREES, A. J. & BRATTIG, N. W. 2001. *Wolbachia* in filarial nematodes: evolutionary aspects and implications for the pathogenesis and treatment of filarial diseases. *Vet Parasitol,* 98**,** 215-38.

BANKS, J. A., NISHIYAMA, T., HASEBE, M., BOWMAN, J. L., GRIBSKOV, M., DEPAMPHILIS, C., ALBERT, V. A., AONO, N., AOYAMA, T., AMBROSE, B. A., ASHTON, N. W., AXTELL, M. J., BARKER, E., BARKER, M. S., BENNETZEN, J. L., BONAWITZ, N. D., CHAPPLE, C., CHENG, C., CORREA, L. G., DACRE, M., DEBARRY, J., DREYER, I., ELIAS, M., ENGSTROM, E. M., ESTELLE, M., FENG, L., FINET, C., FLOYD, S. K., FROMMER, W. B., FUJITA, T., GRAMZOW, L., GUTENSOHN, M., HARHOLT, J., HATTORI, M., HEYL, A., HIRAI, T., HIWATASHI, Y., ISHIKAWA, M., IWATA, M., KAROL, K. G., KOEHLER, B., KOLUKISAOGLU, U., KUBO, M., KURATA, T., LALONDE, S., LI, K., LI, Y., LITT, A., LYONS, E., MANNING, G., MARUYAMA, T., MICHAEL, T. P., MIKAMI, K., MIYAZAKI, S., MORINAGA, S., MURATA, T., MUELLER-ROEBER, B., NELSON, D. R., OBARA, M., OGURI, Y., OLMSTEAD, R. G., ONODERA, N., PETERSEN, B. L., PILS, B., PRIGGE, M., RENSING, S. A., RIANO-PACHON, D. M., ROBERTS, A. W., SATO, Y., SCHELLER, H. V., SCHULZ, B., SCHULZ, C., SHAKIROV, E. V., SHIBAGAKI, N., SHINOHARA, N., SHIPPEN, D. E., SORENSEN, I., SOTOOKA, R., SUGIMOTO, N., SUGITA, M., SUMIKAWA, N., TANURDZIC, M., THEISSEN, G., ULVSKOV, P., WAKAZUKI, S., WENG, J. K., WILLATS, W. W., WIPF, D., WOLF, P. G., YANG, L., ZIMMER, A. D., ZHU, Q., MITROS, T., HELLSTEN, U., LOQUE, D., OTILLAR, R., SALAMOV, A., SCHMUTZ, J., SHAPIRO, H., LINDQUIST, E., et al. 2011. The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science,* 332**,** 960-3.

BAUER, S., VASU, P., PERSSON, S., MORT, A. J. & SOMERVILLE, C. R. 2006. Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. *Proc Natl Acad Sci U S A,* 103**,** 11417-22.

BECKMANN, J. F. & FALLON, A. M. 2013. Detection of the *Wolbachia* protein WPIP0282 in mosquito spermathecae: implications for cytoplasmic incompatibility. *Insect Biochem Mol Biol,* 43**,** 867-78.

BECKMANN, J. F., MARKOWSKI, T. W., WITTHUHN, B. A. & FALLON, A. M. 2013. Detection of the *Wolbachia*-encoded DNA binding protein, HU beta, in mosquito gonads. *Insect Biochem Mol Biol,* 43**,** 272-9.

BEIKO, R. G., HARLOW, T. J. & RAGAN, M. A. 2005. Highways of gene sharing in prokaryotes. *Proc Natl Acad Sci U S A,* 102**,** 14332-7.

BERRY, J., RAJAURE, M., PANG, T. & YOUNG, R. 2012. The spanin complex is essential for lambda lysis. *J Bacteriol,* 194**,** 5667-74.

BERRY, J. D., RAJAURE, M. & YOUNG, R. 2013. Spanin function requires subunit homodimerization through intermolecular disulfide bonds. *Mol Microbiol,* 88**,** 35-47.

BILISKE, J. A., BATISTA, P. D., GRANT, C. L. & HARRIS, H. L. 2011. The bacteriophage WORiC is the active phage element in wRi of Drosophila simulans and represents a conserved class of WO phages. *BMC Microbiol,* 11**,** 251.

BORDENSTEIN, S. R. 2007. Evolutionary genomics: transdomain gene transfers. *Curr Biol,* 17**,** R935-6.

BORDENSTEIN, S. R. & BORDENSTEIN, S. R. 2011. Temperature affects the tripartite interactions between bacteriophage WO, *Wolbachia*, and cytoplasmic incompatibility. *PLoS One,* 6**,** e29106.

BORDENSTEIN, S. R., MARSHALL, M. L., FRY, A. J., KIM, U. & WERNEGREEN, J. J. 2006. The tripartite associations between bacteriophage, *Wolbachia*, and arthropods. *PLoS Pathog,* 2**,** e43.

BORDENSTEIN, S. R., O'HARA, F. P. & WERREN, J. H. 2001. *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. *Nature,* 409**,** 707-10.

BORDENSTEIN, S. R. & REZNIKOFF, W. S. 2005. Mobile DNA in obligate intracellular bacteria. *Nat Rev Microbiol,* 3**,** 688-99.

BORDENSTEIN, S. R. & WERNEGREEN, J. J. 2004. Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Mol Biol Evol,* 21**,** 1981-91.

BORYSOWSKI, J. & GORSKI, A. 2010. Fusion to cell-penetrating peptides will enable lytic enzymes to kill intracellular bacteria. *Med Hypotheses,* 74**,** 164-6.

BOSSAN, B., KOEHNCKE, A. & HAMMERSTEIN, P. 2011. A new model and method for understanding *Wolbachia*-induced cytoplasmic incompatibility. *PLoS One,* 6**,** e19757.

BOTSTEIN, D. 1980. A theory of modular evolution for bacteriophages. *Ann N Y Acad Sci,* 354**,** 484-90.

BOUCHER, H. W., TALBOT, G. H., BRADLEY, J. S., EDWARDS, J. E., GILBERT, D., RICE, L. B., SCHELD, M., SPELLBERG, B. & BARTLETT, J. 2009. Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin Infect Dis,* 48**,** 1-12.

BOURTZIS, K., DOBSON, S. L., XI, Z., RASGON, J. L., CALVITTI, M., MOREIRA, L. A., BOSSIN, H. C., MORETTI, R., BATON, L. A., HUGHES, G. L., MAVINGUI, P. & GILLES, J. R. 2014. Harnessing mosquito-*Wolbachia* symbiosis for vector and disease control. *Acta Trop,* 132 Suppl**,** S150-63.

BOYD, E. F. 2012. Bacteriophage-encoded bacterial virulence factors and phage-pathogenicity island interactions. *Adv Virus Res,* 82**,** 91-118.

BOYD, E. F. & BRUSSOW, H. 2002. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends Microbiol,* 10**,** 521-9.

BRATKE, K. A. & MCLYSAGHT, A. 2008. Identification of multiple independent horizontal gene transfers into poxviruses using a comparative genomics approach. *BMC Evol Biol,* 8**,** 67.

BREEUWER, J. A. & WERREN, J. H. 1993. Cytoplasmic incompatibility and bacterial density in *Nasonia vitripennis*. *Genetics,* 135**,** 565-74.

BRELSFOARD, C. L. & DOBSON, S. L. 2009. *Wolbachia*-based strategies to control insect pests and disease vectors. *AsPac J. Mol. Biol. Biotechnol.,* 17**,** 55-63.

BROCHIER-ARMANET, C., FORTERRE, P. & GRIBALDO, S. 2011. Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr Opin Microbiol,* 14**,** 274-81.

BROWN, J. R. 2003. Ancient horizontal gene transfer. *Nat Rev Genet,* 4**,** 121-32.

BRUSSOW, H., CANCHAYA, C. & HARDT, W. D. 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev,* 68**,** 560-602, table of contents.

BRUTO, B., PRIGENT-COMBARET, C., LUIS, P., HOFF, G., MOËNNE-LOCCOZ, Y. & MULLER, D. 2013. Horizontal Acquisition of Prokaryotic Genes for Eukaryote Functioning and Niche Adaptation. *In:* PONTAROTTI, P. (ed.) *Evolutionary Biology: Exobiology and Evolutionary Mechanisms.* Springer Berlin Heidelberg.

BULGARELLI, D., SCHLAEPPI, K., SPAEPEN, S., VER LOREN VAN THEMAAT, E. & SCHULZE-LEFERT, P. 2013. Structure and functions of the bacterial microbiota of plants. *Annu Rev Plant Biol,* 64**,** 807-38.

BULL, J. J. & TURELLI, M. 2013. *Wolbachia* versus dengue: Evolutionary forecasts. *Evol Med Public Health,* 2013**,** 197-207.

BUSCHIAZZO, E., RITLAND, C., BOHLMANN, J. & RITLAND, K. 2012. Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. *BMC Evol Biol,* 12**,** 8.

CANTAREL, B. L., COUTINHO, P. M., RANCUREL, C., BERNARD, T., LOMBARD, V. & HENRISSAT, B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res,* 37**,** D233-8.

CARVER, T., THOMSON, N., BLEASBY, A., BERRIMAN, M. & PARKHILL, J. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics,* 25**,** 119-20.

CASADESUS, J. & D'ARI, R. 2002. Memory in bacteria and phage. *Bioessays,* 24**,** 512-8.

CASADESUS, J. & LOW, D. 2006. Epigenetic gene regulation in the bacterial world. *Microbiol Mol Biol Rev,* 70**,** 830-56.

CASADESUS, J. & LOW, D. A. 2013. Programmed heterogeneity: epigenetic mechanisms in bacteria. *J Biol Chem,* 288**,** 13929-35.

CASADEVALL, A. 2008. Evolution of intracellular pathogens. *Annu Rev Microbiol,* 62**,** 19-33.

CASIRAGHI, M., BORDENSTEIN, S. R., BALDO, L., LO, N., BENINATI, T., WERNEGREEN, J. J., WERREN, J. H. & BANDI, C. 2005. Phylogeny of *Wolbachia pipientis* based on gltA, groEL and ftsZ gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the *Wolbachia* tree. *Microbiology,* 151**,** 4015-22.

CHAFEE, M. E., FUNK, D. J., HARRISON, R. G. & BORDENSTEIN, S. R. 2010. Lateral phage transfer in obligate intracellular bacteria (*wolbachia*): verification from natural populations. *Mol Biol Evol,* 27**,** 501-5.

CHAFEE, M. E., ZECHER, C. N., GOURLEY, M. L., SCHMIDT, V. T., CHEN, J. H., BORDENSTEIN, S. R., CLARK, M. E. & BORDENSTEIN, S. R. 2011. Decoupling of host-symbiont-phage coadaptations following transfer between insect species. *Genetics,* 187**,** 203-15.

CHEN, Y. & VARANI, G. 2005. Protein families and RNA recognition. *FEBS J,* 272**,** 2088-97.
CHO, K. O., KIM, G. W. & LEE, O. K. 2011. *Wolbachia* bacteria reside in host Golgi-related vesicles whose position is regulated by polarity proteins. *PLoS One,* 6**,** e22703.
CHROSTEK, E., MARIALVA, M. S., ESTEVES, S. S., WEINERT, L. A., MARTINEZ, J., JIGGINS, F. M. & TEIXEIRA, L. 2013. *Wolbachia* variants induce differential protection to viruses in *Drosophila melanogaster*: a phenotypic and phylogenomic analysis. *PLoS Genet,* 9**,** e1003896.
CLARDY, J., FISCHBACH, M. A. & CURRIE, C. R. 2009. The natural history of antibiotics. *Curr Biol,* 19**,** R437-41.
CLARK, M. E., HEATH, B. D., ANDERSON, C. L. & KARR, T. L. 2006. Induced paternal effects mimic cytoplasmic incompatibility in *Drosophila*. *Genetics,* 173**,** 727-34.
CLOKIE, M. R., MILLARD, A. D., LETAROV, A. V. & HEAPHY, S. 2011. Phages in nature. *Bacteriophage,* 1**,** 31-45.
COLLINS, S. R., WELDON, C. W., BANOS, C. & TAYLOR, P. W. 2008. Effects of irradiation dose rate on quality and sterility of Queensland fruit flies, *Bactrocera tryoni* (Froggatt). *J Applied Entomology,* 132**,** 398-405.
CORDAUX, R., BOUCHON, D. & GREVE, P. 2011. The impact of endosymbionts on the evolution of host sex-determination mechanisms. *Trends Genet,* 27**,** 332-41.
COULTHURST, S. J. 2013. The Type VI secretion system - a widespread and versatile cell targeting system. *Res Microbiol,* 164**,** 640-54.
CUBENAS-POTTS, C. & MATUNIS, M. J. 2013. SUMO: a multifaceted modifier of chromatin structure and function. *Dev Cell,* 24**,** 1-12.
DANCHIN, E. G., ROSSO, M. N., VIEIRA, P., DE ALMEIDA-ENGLER, J., COUTINHO, P. M., HENRISSAT, B. & ABAD, P. 2010. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc Natl Acad Sci U S A,* 107**,** 17651-6.
DARBY, A. C., ARMSTRONG, S. D., BAH, G. S., KAUR, G., HUGHES, M. A., KAY, S. M., KOLDKJAER, P., RAINBOW, L., RADFORD, A. D., BLAXTER, M. L., TANYA, V. N., TREES, A. J., CORDAUX, R., WASTLING, J. M. & MAKEPEACE, B. L. 2012. Analysis of gene expression from the *Wolbachia* genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. *Genome Res,* 22**,** 2467-77.
DARLING, A. E., MAU, B. & PERNA, N. T. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One,* 5**,** e11147.
DARRIBA, D., TABOADA, G. L., DOALLO, R. & POSADA, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods,* 9**,** 772.
DEAN, P. 2011. Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *FEMS Microbiol Rev,* 35**,** 1100-25.
DEGNAN, P. H. & MORAN, N. A. 2008. Diverse phage-encoded toxins in a protective insect endosymbiont. *Appl Environ Microbiol,* 74**,** 6782-91.
DHILLON, M. K., SINGH, R., NARESH, J. S. & SHARMA, H. C. 2005. The melon fruit fly, *Bactrocera cucurbitae*: a review of its biology and management. *J Insect Sci,* 5**,** 40.

DODSON, B. L., HUGHES, G. L., PAUL, O., MATACCHIERO, A. C., KRAMER, L. D. & RASGON, J. L. 2014. *Wolbachia* enhances West Nile virus (WNV) infection in the mosquito *Culex tarsalis*. *PLoS Negl Trop Dis,* 8**,** e2965.

DOMENECH, M., GARCIA, E. & MOSCOSO, M. 2011. In vitro destruction of Streptococcus pneumoniae biofilms with bacterial and phage peptidoglycan hydrolases. *Antimicrob Agents Chemother,* 55**,** 4144-8.

DUNNING HOTOPP, J. C. 2011. Horizontal gene transfer between bacteria and animals. *Trends in Genetics,* 27**,** 157-163.

DUNNING HOTOPP, J. C., CLARK, M. E., OLIVEIRA, D. C., FOSTER, J. M., FISCHER, P., MUNOZ TORRES, M. C., GIEBEL, J. D., KUMAR, N., ISHMAEL, N., WANG, S., INGRAM, J., NENE, R. V., SHEPARD, J., TOMKINS, J., RICHARDS, S., SPIRO, D. J., GHEDIN, E., SLATKO, B. E., TETTELIN, H. & WERREN, J. H. 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science,* 317**,** 1753-6.

DUTTON, T. J. & SINKINS, S. P. 2004. Strain-specific quantification of *Wolbachia* density in *Aedes albopictus* and effects of larval rearing conditions. *Insect Mol Biol,* 13**,** 317-22.

DYCK, V. A., HENDRICHS, J. & ROBINSON, A. S. 2005. *Sterile insect technique : principles and practice in area-wide integrated pest management,* Dordrecht, Netherlands, Springer.

DYER, K. A., MINHAS, M. S. & JAENIKE, J. 2005. Expression and modulation of embryonic male-killing in *Drosophila innubila*: opportunities for multilevel selection. *Evolution,* 59**,** 838-48.

ECKBURG, P. B., LEPP, P. W. & RELMAN, D. A. 2003. Archaea and their potential role in human disease. *Infect Immun,* 71**,** 591-6.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res,* 32**,** 1792-7.

ENGELSTADTER, J. & HURST, G. D. D. 2009. The Ecology and Evolution of Microbes that Manipulate Host Reproduction. *Annual Review of Ecology Evolution and Systematics,* 40**,** 127-149.

ENGELSTADTER, J. & TELSCHOW, A. 2009. Cytoplasmic incompatibility and host population structure. *Heredity (Edinb),* 103**,** 196-207.

ENTENZA, J. M., LOEFFLER, J. M., GRANDGIRARD, D., FISCHETTI, V. A. & MOREILLON, P. 2005. Therapeutic effects of bacteriophage Cpl-1 lysin against *Streptococcus pneumoniae* endocarditis in rats. *Antimicrob Agents Chemother,* 49**,** 4789-92.

ESAKOVA, O. & KRASILNIKOV, A. S. 2010. Of proteins and RNA: the RNase P/MRP family. *RNA,* 16**,** 1725-47.

FANG, G., MUNERA, D., FRIEDMAN, D. I., MANDLIK, A., CHAO, M. C., BANERJEE, O., FENG, Z., LOSIC, B., MAHAJAN, M. C., JABADO, O. J., DEIKUS, G., CLARK, T. A., LUONG, K., MURRAY, I. A., DAVIS, B. M., KEREN-PAZ, A., CHESS, A., ROBERTS, R. J., KORLACH, J., TURNER, S. W., KUMAR, V., WALDOR, M. K. & SCHADT, E. E. 2012. Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat Biotechnol,* 30**,** 1232-9.

FAST, E. M., TOOMEY, M. E., PANARAM, K., DESJARDINS, D., KOLACZYK, E. D. & FRYDMAN, H. M. 2011. *Wolbachia* enhance *Drosophila* stem cell proliferation and target the germline stem cell niche. *Science,* 334**,** 990-2.

FASTREZ, J. 1996. Phage lysozymes. *In:* JOLLES, P. (ed.) *Lysozymes: Model Enzymes in Biochemistry and Biology.* Birkhauser Verlag.

FENN, K. & BLAXTER, M. 2004. Are filarial nematode *Wolbachia* obligate mutualist symbionts? *Trends Ecol Evol,* 19**,** 163-6.

FENN, K., CONLON, C., JONES, M., QUAIL, M. A., HOLROYD, N. E., PARKHILL, J. & BLAXTER, M. 2006. Phylogenetic relationships of the *Wolbachia* of nematodes and arthropods. *PLoS Pathog,* 2**,** e94.

FENTON, M., CASEY, P. G., HILL, C., GAHAN, C. G., ROSS, R. P., MCAULIFFE, O., O'MAHONY, J., MAHER, F. & COFFEY, A. 2010. The truncated phage lysin CHAP(k) eliminates *Staphylococcus aureus* in the nares of mice. *Bioengineered bugs,* 1**,** 404-407.

FERGUSON, S. B., BLUNDON, M. A., KLOVSTAD, M. S. & SCHUPBACH, T. 2012. Modulation of gurken translation by insulin and TOR signaling in *Drosophila*. *J Cell Sci,* 125**,** 1407-19.

FERRI, E., BAIN, O., BARBUTO, M., MARTIN, C., LO, N., UNI, S., LANDMANN, F., BACCEI, S. G., GUERRERO, R., DE SOUZA LIMA, S., BANDI, C., WANJI, S., DIAGNE, M. & CASIRAGHI, M. 2011. New insights into the evolution of *Wolbachia* infections in filarial nematodes inferred from a large range of screened species. *PLoS One,* 6**,** e20843.

FISCHETTI, V. A. 2010. Bacteriophage endolysins: a novel anti-infective to control Gram-positive pathogens. *Int J Med Microbiol,* 300**,** 357-62.

FLORES, G. E., WAGNER, I. D., LIU, Y. & REYSENBACH, A. L. 2012. Distribution, abundance, and diversity patterns of the thermoacidophilic "deep-sea hydrothermal vent euryarchaeota 2". *Front Microbiol,* 3**,** 47.

FOSTER, J., GANATRA, M., KAMAL, I., WARE, J., MAKAROVA, K., IVANOVA, N., BHATTACHARYYA, A., KAPATRAL, V., KUMAR, S., POSFAI, J., VINCZE, T., INGRAM, J., MORAN, L., LAPIDUS, A., OMELCHENKO, M., KYRPIDES, N., GHEDIN, E., WANG, S., GOLTSMAN, E., JOUKOV, V., OSTROVSKAYA, O., TSUKERMAN, K., MAZUR, M., COMB, D., KOONIN, E. & SLATKO, B. 2005. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol,* 3**,** e121.

FRANCESCONI, F. & LUPI, O. 2012. Myiasis. *Clin Microbiol Rev,* 25**,** 79-105.

FUJII, Y., KAGEYAMA, D., HOSHIZAKI, S., ISHIKAWA, H. & SASAKI, T. 2001. Transfection of *Wolbachia* in Lepidoptera: the feminizer of the adzuki bean borer *Ostrinia scapulalis* causes male killing in the Mediterranean flour moth *Ephestia kuehniella*. *Proc Biol Sci,* 268**,** 855-9.

FUJII, Y., KUBO, T., ISHIKAWA, H. & SASAKI, T. 2004. Isolation and characterization of the bacteriophage WO from *Wolbachia*, an arthropod endosymbiont. *Biochem Biophys Res Commun,* 317**,** 1183-8.

FURTADO, A. R., ESSID, M., PERRINET, S., BALANA, M. E., YODER, N., DEHOUX, P. & SUBTIL, A. 2013. The chlamydial OTU domain-containing protein ChlaOTU is an early type III secretion effector targeting ubiquitin and NDP52. *Cell Microbiol,* 15**,** 2064-79.

GAGLIA, M. M., COVARRUBIAS, S., WONG, W. & GLAUNSINGER, B. A. 2012. A common strategy for host RNA degradation by divergent viruses. *J Virol,* 86**,** 9527-30.

GAMSTON, C. & RASGON, J. 2007. Maintaining *Wolbachia* in cell-free medium. *J Vis Exp***,** 223.

GAVOTTE, L., HENRI, H., STOUTHAMER, R., CHARIF, D., CHARLAT, S., BOULETREAU, M. & VAVRE, F. 2007. A Survey of the bacteriophage WO in the endosymbiotic bacteria *Wolbachia*. *Mol Biol Evol,* 24**,** 427-35.

GLADYSHEV, E. A., MESELSON, M. & ARKHIPOVA, I. R. 2008. Massive horizontal gene transfer in bdelloid rotifers. *Science,* 320**,** 1210-3.

GOMEZ-VALERO, L., SORIANO-NAVARRO, M., PEREZ-BROCAL, V., HEDDI, A., MOYA, A., GARCIA-VERDUGO, J. M. & LATORRE, A. 2004. Coexistence of *Wolbachia* with *Buchnera aphidicola* and a secondary symbiont in the aphid *Cinara cedri*. *J Bacteriol,* 186**,** 6626-33.

GROTH, A. C., FISH, M., NUSSE, R. & CALOS, M. P. 2004. Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. *Genetics,* 166**,** 1775-82.

GUEGUEN, G., ONEMOLA, B. & GOVIND, S. 2012. Association of a new *Wolbachia* strain with, and its effects on, *Leptopilina victoriae*, a virulent wasp parasitic to *Drosophila spp*. *Appl Environ Microbiol,* 78**,** 5962-6.

GUERRIER-TAKADA, C. & ALTMAN, S. 1984. Catalytic activity of an RNA molecule prepared by transcription in vitro. *Science,* 223**,** 285-6.

GUINDON, S., DUFAYARD, J. F., LEFORT, V., ANISIMOVA, M., HORDIJK, W. & GASCUEL, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol,* 59**,** 307-21.

HALGASOVA, N., UGORCAKOVA, J., GEROVA, M., TIMKO, J. & BUKOVSKA, G. 2010. Isolation and characterization of bacteriophage PhiBP from *Paenibacillus polymyxa* CCM 7400. *FEMS Microbiol Lett,* 305**,** 128-35.

HEATH, B. D., BUTCHER, R. D., WHITFIELD, W. G. & HUBBARD, S. F. 1999. Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. *Curr Biol,* 9**,** 313-6.

HEITHOFF, D. M., SINSHEIMER, R. L., LOW, D. A. & MAHAN, M. J. 1999. An essential role for DNA adenine methylation in bacterial virulence. *Science,* 284**,** 967-70.

HENRICHFREISE, B., SCHIEFER, A., SCHNEIDER, T., NZUKOU, E., POELLINGER, C., HOFFMANN, T. J., JOHNSTON, K. L., MOELLEKEN, K., WIEDEMANN, I., PFARR, K., HOERAUF, A. & SAHL, H. G. 2009. Functional conservation of the lipid II biosynthesis pathway in the cell wall-less bacteria *Chlamydia* and *Wolbachia*: why is lipid II needed? *Mol Microbiol,* 73**,** 913-23.

HERTIG, M. & WOLBACH, S. B. 1924. Studies on Rickettsia-Like Micro-Organisms in Insects. *J Med Res,* 44**,** 329-374 7.

HILGENBOECKER, K., HAMMERSTEIN, P., SCHLATTMANN, P., TELSCHOW, A. & WERREN, J. H. 2008. How many species are infected with *Wolbachia*?--A statistical analysis of current data. *FEMS Microbiol Lett,* 281**,** 215-20.

HOFFMANN, A. A., CLANCY, D. & DUNCAN, J. 1996. Naturally-occurring *Wolbachia* infection in *Drosophila simulans* that does not cause cytoplasmic incompatibility. *Heredity (Edinb),* 76 (Pt 1)**,** 1-8.

HUGHES, G. L., REN, X., RAMIREZ, J. L., SAKAMOTO, J. M., BAILEY, J. A., JEDLICKA, A. E. & RASGON, J. L. 2011. *Wolbachia* infections in *Anopheles gambiae* cells: transcriptomic characterization of a novel host-symbiont interaction. *PLoS Pathog,* 7**,** e1001296.

HURST, G. D. & JIGGINS, F. M. 2000. Male-killing bacteria in insects: mechanisms, incidence, and implications. *Emerg Infect Dis,* 6**,** 329-36.

HUSNIK, F., NIKOH, N., KOGA, R., ROSS, L., DUNCAN, R. P., FUJIE, M., TANAKA, M., SATOH, N., BACHTROG, D., WILSON, A. C., VON DOHLEN, C. D., FUKATSU, T. & MCCUTCHEON, J. P. 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell,* 153**,** 1567-78.

INTERNATIONAL APHID GENOMICS CONSORTIUM 2010. Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol,* 8**,** e1000313.

IOANNIDIS, P., DUNNING HOTOPP, J. C., SAPOUNTZIS, P., SIOZIOS, S., TSIAMIS, G., BORDENSTEIN, S. R., BALDO, L., WERREN, J. H. & BOURTZIS, K. 2007. New criteria for selecting the origin of DNA replication in *Wolbachia* and closely related bacteria. *BMC Genomics,* 8**,** 182.

ISHMAEL, N., DUNNING HOTOPP, J. C., IOANNIDIS, P., BIBER, S., SAKAMOTO, J., SIOZIOS, S., NENE, V., WERREN, J., BOURTZIS, K., BORDENSTEIN, S. R. & TETTELIN, H. 2009. Extensive genomic diversity of closely related *Wolbachia* strains. *Microbiology,* 155**,** 2211-22.

ITURBE-ORMAETXE, I., BURKE, G. R., RIEGLER, M. & O'NEILL, S. L. 2005. Distribution, expression, and motif variability of ankyrin domain genes in *Wolbachia pipientis*. *J Bacteriol,* 187**,** 5136-45.

ITURBE-ORMAETXE, I., WALKER, T. & SL, O. N. 2011. *Wolbachia* and the biological control of mosquito-borne disease. *EMBO Rep,* 12**,** 508-18.

IYER, L. M., ZHANG, D., ROGOZIN, I. B. & ARAVIND, L. 2011. Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res,* 39**,** 9473-97.

JAENIKE, J. 2007. Spontaneous emergence of a new *Wolbachia* phenotype. *Evolution,* 61**,** 2244-52.

JAENIKE, J. 2009. Coupled population dynamics of endosymbionts within and between hosts. *Oikos,* 118**,** 353-362.

JAENIKE, J., DYER, K. A., CORNISH, C. & MINHAS, M. S. 2006. Asymmetrical reinforcement and *Wolbachia* infection in *Drosophila*. *PLoS Biol,* 4**,** e325.

JENKINS, T. G. & CARRELL, D. T. 2012. The sperm epigenome and potential implications for the developing embryo. *Reproduction,* 143**,** 727-34.

KAMBRIS, Z., COOK, P. E., PHUC, H. K. & SINKINS, S. P. 2009. Immune activation by life-shortening *Wolbachia* and reduced filarial competence in mosquitoes. *Science,* 326**,** 134-6.

KATO, S. & WATANABE, K. 2010. Ecological and evolutionary interactions in syntrophic methanogenic consortia. *Microbes Environ,* 25**,** 145-51.

KAY, S., HAHN, S., MAROIS, E., HAUSE, G. & BONAS, U. 2007. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science,* 318**,** 648-51.

KEARSE, M., MOIR, R., WILSON, A., STONES-HAVAS, S., CHEUNG, M., STURROCK, S., BUXTON, S., COOPER, A., MARKOWITZ, S., DURAN, C., THIERER, T., ASHTON, B., MEINTJES, P. & DRUMMOND, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28**,** 1647-9.

KEELING, P. J. & PALMER, J. D. 2008. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet,* 9**,** 605-18.

KELLEY, L. A. & STERNBERG, M. J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc,* 4**,** 363-71.

KENT, B. N. & BORDENSTEIN, S. R. 2010. Phage WO of *Wolbachia*: lambda of the endosymbiont world. *Trends Microbiol,* 18**,** 173-81.

KENT, B. N., FUNKHOUSER, L. J., SETIA, S. & BORDENSTEIN, S. R. 2011a. Evolutionary genomics of a temperate bacteriophage in an obligate intracellular bacteria (*Wolbachia*). *PLoS One,* 6**,** e24984.

KENT, B. N., SALICHOS, L., GIBBONS, J. G., ROKAS, A., NEWTON, I. L., CLARK, M. E. & BORDENSTEIN, S. R. 2011b. Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*) determined by targeted genome capture. *Genome Biol Evol,* 3**,** 209-18.

KLASSON, L. & ANDERSSON, S. G. 2004. Evolution of minimal-gene-sets in host-dependent bacteria. *Trends Microbiol,* 12**,** 37-43.

KLASSON, L., KAMBRIS, Z., COOK, P. E., WALKER, T. & SINKINS, S. P. 2009a. Horizontal gene transfer between *Wolbachia* and the mosquito *Aedes aegypti*. *BMC Genomics,* 10**,** 33.

KLASSON, L., WALKER, T., SEBAIHIA, M., SANDERS, M. J., QUAIL, M. A., LORD, A., SANDERS, S., EARL, J., O'NEILL, S. L., THOMSON, N., SINKINS, S. P. & PARKHILL, J. 2008. Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. *Molecular biology and evolution,* 25**,** 1877-87.

KLASSON, L., WESTBERG, J., SAPOUNTZIS, P., NASLUND, K., LUTNAES, Y., DARBY, A. C., VENETI, Z., CHEN, L., BRAIG, H. R., GARRETT, R., BOURTZIS, K. & ANDERSSON, S. G. 2009b. 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**,** 5725-30.

KNIPLING, E. F. 1959. Sterile-male method of population control. *Science,* 130**,** 902-4.

KONDRASHOV, F. A., KOONIN, E. V., MORGUNOV, I. G., FINOGENOVA, T. V. & KONDRASHOVA, M. N. 2006. Evolution of glyoxylate cycle enzymes in Metazoa: evidence of multiple horizontal transfer events and pseudogene formation. *Biol Direct,* 1**,** 31.

KOONIN, E. V., MAKAROVA, K. S., ROGOZIN, I. B., DAVIDOVIC, L., LETELLIER, M. C. & PELLEGRINI, L. 2003. The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers. *Genome Biol,* 4**,** R19.

KORCZYNSKA, J. E., DANIELSEN, S., SCHAGERLOF, U., TURKENBURG, J. P., DAVIES, G. J., WILSON, K. S. & TAYLOR, E. J. 2010. The structure of a

family GH25 lysozyme from *Aspergillus fumigatus*. *Acta Crystallogr Sect F Struct Biol Cryst Commun,* 66**,** 973-7.
LANDMANN, F., ORSI, G. A., LOPPIN, B. & SULLIVAN, W. 2009. *Wolbachia*-mediated cytoplasmic incompatibility is associated with impaired histone deposition in the male pronucleus. *PLoS Pathog,* 5**,** e1000343.
LE CLEC'H, W., CHEVALIER, F. D., GENTY, L., BERTAUX, J., BOUCHON, D. & SICARD, M. 2013. Cannibalism and predation as paths for horizontal passage of *Wolbachia* between terrestrial isopods. *PLoS One,* 8**,** e60232.
LEFTWICH, P. T., KOUKIDOU, M., REMPOULAKIS, P., GONG, H. F., ZACHAROPOULOU, A., FU, G., CHAPMAN, T., ECONOMOPOULOS, A., VONTAS, J. & ALPHEY, L. 2014. Genetic elimination of field-cage populations of Mediterranean fruit flies. *Proc Biol Sci,* 281.
LENSKI, R. E., ROSE, M. R., SIMPSON, S. C. & TADLER, S. C. 1991. Long-Term Experimental Evolution in *Escherichia-Coli* .1. Adaptation and Divergence during 2,000 Generations. *American Naturalist,* 138**,** 1315-1341.
LEPAGE, D. & BORDENSTEIN, S. R. 2013. *Wolbachia*: Can we save lives with a great pandemic? *Trends Parasitol,* 29**,** 385-93.
LI, S. J. & HOCHSTRASSER, M. 1999. A new protease required for cell-cycle progression in yeast. *Nature,* 398**,** 246-51.
LIBRADO, P. & ROZAS, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* 25**,** 1451-2.
LIU, F. & ALTMAN, S. 1994. Differential evolution of substrates for an RNA enzyme in the presence and absence of its protein cofactor. *Cell,* 77**,** 1093-100.
LIU, J., DEHBI, M., MOECK, G., ARHIN, F., BAUDA, P., BERGERON, D., CALLEJO, M., FERRETTI, V., HA, N., KWAN, T., MCCARTY, J., SRIKUMAR, R., WILLIAMS, D., WU, J. J., GROS, P., PELLETIER, J. & DUBOW, M. 2004. Antimicrobial drug discovery through bacteriophage genomics. *Nat Biotechnol,* 22**,** 185-91.
LOEFFLER, J. M., DJURKOVIC, S. & FISCHETTI, V. A. 2003. Phage lytic enzyme Cpl-1 as a novel antimicrobial for pneumococcal bacteremia. *Infect Immun,* 71**,** 6199-204.
LUNDIN, D., GRIBALDO, S., TORRENTS, E., SJOBERG, B. M. & POOLE, A. M. 2010. Ribonucleotide reduction - horizontal transfer of a required function spans all three domains. *BMC Evol Biol,* 10**,** 383.
MACHIDA, M., ASAI, K., SANO, M., TANAKA, T., KUMAGAI, T., TERAI, G., KUSUMOTO, K., ARIMA, T., AKITA, O., KASHIWAGI, Y., ABE, K., GOMI, K., HORIUCHI, H., KITAMOTO, K., KOBAYASHI, T., TAKEUCHI, M., DENNING, D. W., GALAGAN, J. E., NIERMAN, W. C., YU, J., ARCHER, D. B., BENNETT, J. W., BHATNAGAR, D., CLEVELAND, T. E., FEDOROVA, N. D., GOTOH, O., HORIKAWA, H., HOSOYAMA, A., ICHINOMIYA, M., IGARASHI, R., IWASHITA, K., JUVVADI, P. R., KATO, M., KATO, Y., KIN, T., KOKUBUN, A., MAEDA, H., MAEYAMA, N., MARUYAMA, J., NAGASAKI, H., NAKAJIMA, T., ODA, K., OKADA, K., PAULSEN, I., SAKAMOTO, K., SAWANO, T., TAKAHASHI, M., TAKASE, K., TERABAYASHI, Y., WORTMAN, J. R., YAMADA, O., YAMAGATA, Y., ANAZAWA, H., HATA, Y., KOIDE, Y., KOMORI, T., KOYAMA, Y.,

MINETOKI, T., SUHARNAN, S., TANAKA, A., ISONO, K., KUHARA, S., OGASAWARA, N. & KIKUCHI, H. 2005. Genome sequencing and analysis of *Aspergillus oryzae*. *Nature,* 438**,** 1157-61.

MAKAROVA, K. S., ARAVIND, L. & KOONIN, E. V. 2000. A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and *Chlamydia pneumoniae*. *Trends Biochem Sci,* 25**,** 50-2.

MARTINEZ-FLEITES, C., KORCZYNSKA, J. E., DAVIES, G. J., COPE, M. J., TURKENBURG, J. P. & TAYLOR, E. J. 2009. The crystal structure of a family GH25 lysozyme from *Bacillus anthracis* implies a neighboring-group catalytic mechanism with retention of anomeric configuration. *Carbohydr Res,* 344**,** 1753-7.

MARVIN, M. C. & ENGELKE, D. R. 2009. Broadening the mission of an RNA enzyme. *J Cell Biochem,* 108**,** 1244-51.

MASUI, S., KAMODA, S., SASAKI, T. & ISHIKAWA, H. 2000. Distribution and evolution of bacteriophage WO in *Wolbachia*, the endosymbiont causing sexual alterations in arthropods. *J Mol Evol,* 51**,** 491-7.

MAYORAL, J. G., ETEBARI, K., HUSSAIN, M., KHROMYKH, A. A. & ASGARI, S. 2014. *Wolbachia* infection modifies the profile, shuttling and structure of microRNAs in a mosquito cell line. *PLoS One,* 9**,** e96107.

MCCLURE, M. A. 2001. Evolution of the DUT gene: horizontal transfer between host and pathogen in all three domains of life. *Curr Protein Pept Sci,* 2**,** 313-24.

MCDONALD, T. R., DIETRICH, F. S. & LUTZONI, F. 2012. Multiple horizontal gene transfers of ammonium transporters/ammonia permeases from prokaryotes to eukaryotes: toward a new functional and evolutionary classification. *Mol Biol Evol,* 29**,** 51-60.

MCMENIMAN, C. J., LANE, A. M., FONG, A. W., VORONIN, D. A., ITURBE-ORMAETXE, I., YAMADA, R., MCGRAW, E. A. & O'NEILL, S. L. 2008. Host adaptation of a *Wolbachia* strain after long-term serial passage in mosquito cell lines. *Appl Environ Microbiol,* 74**,** 6963-9.

MCMENIMAN, C. J., LANE, R. V., CASS, B. N., FONG, A. W., SIDHU, M., WANG, Y. F. & O'NEILL, S. L. 2009. Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti*. *Science,* 323**,** 141-4.

MCNULTY, S. N., ABUBUCKER, S., SIMON, G. M., MITREVA, M., MCNULTY, N. P., FISCHER, K., CURTIS, K. C., BRATTIG, N. W., WEIL, G. J. & FISCHER, P. U. 2012. Transcriptomic and proteomic analyses of a *Wolbachia*-free filarial parasite provide evidence of trans-kingdom horizontal gene transfer. *PLoS One,* 7**,** e45777.

METCALF, J. A. & BORDENSTEIN, S. R. 2012. The complexity of virus systems: the case of endosymbionts. *Curr Opin Microbiol,* 15**,** 546-52.

METCALF, J. A., JO, M., BORDENSTEIN, S. R., JAENIKE, J. & BORDENSTEIN, S. R. 2014. Recent genome reduction of *Wolbachia* in *Drosophila recens* targets phage WO and narrows candidates for reproductive parasitism. *PeerJ,* In Press.

MEYER, D. F., NOROY, C., MOUMENE, A., RAFFAELE, S., ALBINA, E. & VACHIERY, N. 2013. Searching algorithm for type IV secretion system effectors 1.0: a tool for predicting type IV effectors and exploring their genomic context. *Nucleic Acids Res,* 41**,** 9218-29.

MILLER, W. J., EHRMAN, L. & SCHNEIDER, D. 2010. Infectious speciation revisited: impact of symbiont-depletion on female fitness and mating behavior of *Drosophila paulistorum*. *PLoS Pathog,* 6**,** e1001214.

MIN, K. T. & BENZER, S. 1997. *Wolbachia*, normally a symbiont of Drosophila, can be virulent, causing degeneration and early death. *Proc Natl Acad Sci U S A,* 94**,** 10792-6.

MIROSHNICHENKO, M. L. & BONCH-OSMOLOVSKAYA, E. A. 2006. Recent developments in the thermophilic microbiology of deep-sea hydrothermal vents. *Extremophiles,* 10**,** 85-96.

MORAN, N. A., MCCUTCHEON, J. P. & NAKABACHI, A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet,* 42**,** 165-90.

MORAN, Y., FREDMAN, D., SZCZESNY, P., GRYNBERG, M. & TECHNAU, U. 2012. Recurrent Horizontal Transfer of Bacterial Toxin Genes to Eukaryotes. *Mol Biol Evol*.

MOREIRA, L. A., ITURBE-ORMAETXE, I., JEFFERY, J. A., LU, G., PYKE, A. T., HEDGES, L. M., ROCHA, B. C., HALL-MENDELIN, S., DAY, A., RIEGLER, M., HUGO, L. E., JOHNSON, K. N., KAY, B. H., MCGRAW, E. A., VAN DEN HURK, A. F., RYAN, P. A. & O'NEILL, S. L. 2009. A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, Chikungunya, and *Plasmodium*. *Cell,* 139**,** 1268-78.

MORENO-HAGELSIEB, G. & LATIMER, K. 2008. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics,* 24**,** 319-24.

NEGRI, I., FRANCHINI, A., GONELLA, E., DAFFONCHIO, D., MAZZOGLIO, P. J., MANDRIOLI, M. & ALMA, A. 2009. Unravelling the *Wolbachia* evolutionary role: the reprogramming of the host genomic imprinting. *Proc Biol Sci,* 276**,** 2485-91.

NELSON, K. E., CLAYTON, R. A., GILL, S. R., GWINN, M. L., DODSON, R. J., HAFT, D. H., HICKEY, E. K., PETERSON, J. D., NELSON, W. C., KETCHUM, K. A., MCDONALD, L., UTTERBACK, T. R., MALEK, J. A., LINHER, K. D., GARRETT, M. M., STEWART, A. M., COTTON, M. D., PRATT, M. S., PHILLIPS, C. A., RICHARDSON, D., HEIDELBERG, J., SUTTON, G. G., FLEISCHMANN, R. D., EISEN, J. A., WHITE, O., SALZBERG, S. L., SMITH, H. O., VENTER, J. C. & FRASER, C. M. 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature,* 399**,** 323-9.

NEVALAINEN, T. J., GRAHAM, G. G. & SCOTT, K. F. 2008. Antibacterial actions of secreted phospholipases A2. Review. *Biochim Biophys Acta,* 1781**,** 1-9.

NEWTON, I. L. & BORDENSTEIN, S. R. 2011. Correlations between bacterial ecology and mobile DNA. *Curr Microbiol,* 62**,** 198-208.

NIKOH, N., MCCUTCHEON, J. P., KUDO, T., MIYAGISHIMA, S. Y., MORAN, N. A. & NAKABACHI, A. 2010. Bacterial genes in the aphid genome: absence of functional gene transfer from *Buchnera* to its host. *PLoS Genet,* 6**,** e1000827.

NIKOH, N., TANAKA, K., SHIBATA, F., KONDO, N., HIZUME, M., SHIMADA, M. & FUKATSU, T. 2008. *Wolbachia* genome integrated in an insect chromosome: evolution and fate of laterally transferred endosymbiont genes. *Genome Res,* 18**,** 272-80.

NORIKUNI, K., DAI, H. & TSUGUO, K. 2008. Effect of irradiation on mating performance and mating ability in the West Indian sweetpotato weevil, *Euscepes postfasciatus*. *Entomologia Experimentalis et Applicata,* 127**,** 229-336(8).

O'CONNOR, E. M. & SHAND, R. F. 2002. Halocins and sulfolobicins: the emerging story of archaeal protein and peptide antibiotics. *J Ind Microbiol Biotechnol,* 28**,** 23-31.

O'CONNOR, L., PLICHART, C., SANG, A. C., BRELSFOARD, C. L., BOSSIN, H. C. & DOBSON, S. L. 2012. Open release of male mosquitoes infected with a *wolbachia* biopesticide: field performance and infection containment. *PLoS Negl Trop Dis,* 6**,** e1797.

OLIVER, K. M., DEGNAN, P. H., HUNTER, M. S. & MORAN, N. A. 2009. Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science,* 325**,** 992-4.

ORCUTT, B. N., BACH, W., BECKER, K., FISHER, A. T., HENTSCHER, M., TONER, B. M., WHEAT, C. G. & EDWARDS, K. J. 2011a. Colonization of subsurface microbial observatories deployed in young ocean crust. *ISME J,* 5**,** 692-703.

ORCUTT, B. N., SYLVAN, J. B., KNAB, N. J. & EDWARDS, K. J. 2011b. Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiol Mol Biol Rev,* 75**,** 361-422.

OREN, A. 2002. Molecular ecology of extremely halophilic Archaea and Bacteria. *FEMS Microbiol Ecol,* 39**,** 1-7.

OSBORNE, T. B. & CAMPBELL, G. F. 1896. The proteids of the potato. *Journal of the American Chemical Society,* 18**,** 575-582.

OSEI-AMO, S., HUSSAIN, M., O'NEILL, S. L. & ASGARI, S. 2012. *Wolbachia*-induced aae-miR-12 miRNA negatively regulates the expression of MCT1 and MCM6 genes in *Wolbachia*-infected mosquito cell line. *PLoS One,* 7**,** e50049.

PANDEY, D. P. & GERDES, K. 2005. Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res,* 33**,** 966-76.

PECK-MILLER, K. A. & ALTMAN, S. 1991. Kinetics of the processing of the precursor to 4.5 S RNA, a naturally occurring substrate for RNase P from *Escherichia coli*. *J Mol Biol,* 221**,** 1-5.

PICHON, S., BOUCHON, D., CORDAUX, R., CHEN, L., GARRETT, R. A. & GREVE, P. 2009. Conservation of the Type IV secretion system throughout *Wolbachia* evolution. *Biochem Biophys Res Commun,* 385**,** 557-62.

PICHON, S., BOUCHON, D., LIU, C., CHEN, L., GARRETT, R. A. & GREVE, P. 2012. The expression of one ankyrin pk2 allele of the WO prophage is correlated with the *Wolbachia* feminizing effect in isopods. *BMC Microbiol,* 12**,** 55.

PINTO, S. B., STAINTON, K., HARRIS, S., KAMBRIS, Z., SUTTON, E. R., BONSALL, M. B., PARKHILL, J. & SINKINS, S. P. 2013. Transcriptional Regulation of *Culex pipiens* Mosquitoes by *Wolbachia* Influences Cytoplasmic Incompatibility. *PLoS Pathog,* 9**,** e1003647.

POINSOT, D., CHARLAT, S. & MERCOT, H. 2003. On the mechanism of *Wolbachia*-induced cytoplasmic incompatibility: confronting the models with the facts. *Bioessays,* 25**,** 259-65.

POPA, O. & DAGAN, T. 2011. Trends and barriers to lateral gene transfer in prokaryotes. *Curr Opin Microbiol,* 14**,** 615-23.

PUIGBO, P., WOLF, Y. I. & KOONIN, E. V. 2009. Search for a 'Tree of Life' in the thicket of the phylogenetic forest. *J Biol,* 8**,** 59.

QUAST, C., PRUESSE, E., YILMAZ, P., GERKEN, J., SCHWEER, T., YARZA, P., PEPLIES, J. & GLOCKNER, F. O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res,* 41**,** D590-6.

RANCES, E., VORONIN, D., TRAN-VAN, V. & MAVINGUI, P. 2008. Genetic and functional characterization of the type IV secretion system in *Wolbachia*. *J Bacteriol,* 190**,** 5020-30.

RASGON, J. L., GAMSTON, C. E. & REN, X. 2006. Survival of *Wolbachia pipientis* in cell-free medium. *Appl Environ Microbiol,* 72**,** 6934-7.

RAYCHOUDHURY, R., BALDO, L., OLIVEIRA, D. C. & WERREN, J. H. 2009. Modes of acquisition of *Wolbachia*: horizontal transfer, hybrid introgression, and codivergence in the *Nasonia* species complex. *Evolution,* 63**,** 165-83.

REISENAUER, A., KAHNG, L. S., MCCOLLUM, S. & SHAPIRO, L. 1999. Bacterial DNA methylation: a cell cycle regulator? *J Bacteriol,* 181**,** 5135-9.

REYNOLDS, K. T. & HOFFMANN, A. A. 2002. Male age, host effects and the weak expression or non-expression of cytoplasmic incompatibility in *Drosophila* strains infected by maternally transmitted *Wolbachia*. *Genet Res,* 80**,** 79-87.

REYNOLDS, S. & ROLFF, J. 2008. Immune function keeps endosymbionts under control. *J Biol,* 7**,** 28.

REYSENBACH, A. L., LIU, Y., BANTA, A. B., BEVERIDGE, T. J., KIRSHTEIN, J. D., SCHOUTEN, S., TIVEY, M. K., VON DAMM, K. L. & VOYTEK, M. A. 2006. A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. *Nature,* 442**,** 444-7.

REYSENBACH, A. L., LIU, Y., LINDGREN, A. R., WAGNER, I. D., SISLAK, C. D., METS, A. & SCHOUTEN, S. 2013. *Mesoaciditoga lauensis* gen. nov., sp. nov., a moderate thermoacidophilic Thermotogales from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol*.

RICHARDSON, M. F., WEINERT, L. A., WELCH, J. J., LINHEIRO, R. S., MAGWIRE, M. M., JIGGINS, F. M. & BERGMAN, C. M. 2012. Population genomics of the *Wolbachia* endosymbiont in *Drosophila melanogaster*. *PLoS Genet,* 8**,** e1003129.

RIEGLER, M., CHARLAT, S., STAUFFER, C. & MERCOT, H. 2004. *Wolbachia* transfer from *Rhagoletis cerasi* to *Drosophila simulans*: investigating the outcomes of host-symbiont coevolution. *Appl Environ Microbiol,* 70**,** 273-9.

RIGAUD, T., PENNINGS, P. S. & JUCHAULT, P. 2001. *Wolbachia* bacteria effects after experimental interspecific transfers in terrestrial isopods. *J Invertebr Pathol,* 77**,** 251-7.

RIKIHISA, Y. & LIN, M. 2010. *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* type IV secretion and Ank proteins. *Curr Opin Microbiol,* 13**,** 59-66.

RILEY, D. R., ANGIUOLI, S. V., CRABTREE, J., DUNNING HOTOPP, J. C. & TETTELIN, H. 2012. Using Sybil for interactive comparative genomics of microbes on the web. *Bioinformatics,* 28**,** 160-6.

RONQUIST, F., TESLENKO, M., VAN DER MARK, P., AYRES, D. L., DARLING, A., HOHNA, S., LARGET, B., LIU, L., SUCHARD, M. A. & HUELSENBECK, J. P.

2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol,* 61**,** 539-42.

ROSS, C. L., DYER, K. A., EREZ, T., MILLER, S. J., JAENIKE, J. & MARKOW, T. A. 2003. Rapid divergence of microsatellite abundance among species of *Drosophila*. *Mol Biol Evol,* 20**,** 1143-57.

RUPP, J., SOLBACH, W. & GIEFFERS, J. 2007. Prevalence, genetic conservation and transmissibility of the *Chlamydia pneumoniae* bacteriophage (phiCpn1). *FEMS Microbiol Lett,* 273**,** 45-9.

SANOGO, Y. O. & DOBSON, S. L. 2006. WO bacteriophage transcription in *Wolbachia*-infected *Culex pipiens*. *Insect Biochem Mol Biol,* 36**,** 80-5.

SANOGO, Y. O., EITAM, A. & DOBSON, S. L. 2005. No evidence for bacteriophage WO orf7 correlation with *Wolbachia*-induced cytoplasmic incompatibility in the *Culex pipien*s complex (Culicidae: Diptera). *J Med Entomol,* 42**,** 789-94.

SARAOGI, I. & SHAN, S. O. 2014. Co-translational protein targeting to the bacterial membrane. *Biochim Biophys Acta,* 1843**,** 1433-41.

SARIDAKI, A., SAPOUNTZIS, P., HARRIS, H. L., BATISTA, P. D., BILISKE, J. A., PAVLIKAKI, H., OEHLER, S., SAVAKIS, C., BRAIG, H. R. & BOURTZIS, K. 2011. *Wolbachia* prophage DNA adenine methyltransferase genes in different *Drosophila*-*Wolbachia* associations. *PLoS One,* 6**,** e19708.

SAVVA, C. G., DEWEY, J. S., DEATON, J., WHITE, R. L., STRUCK, D. K., HOLZENBURG, A. & YOUNG, R. 2008. The holin of bacteriophage lambda forms rings with large diameter. *Mol Microbiol,* 69**,** 784-793.

SCHERER, G. F., RYU, S. B., WANG, X., MATOS, A. R. & HEITZ, T. 2010. Patatin-related phospholipase A: nomenclature, subfamilies and functions in plants. *Trends Plant Sci,* 15**,** 693-700.

SCHONKNECHT, G., CHEN, W. H., TERNES, C. M., BARBIER, G. G., SHRESTHA, R. P., STANKE, M., BRAUTIGAM, A., BAKER, B. J., BANFIELD, J. F., GARAVITO, R. M., CARR, K., WILKERSON, C., RENSING, S. A., GAGNEUL, D., DICKENSON, N. E., OESTERHELT, C., LERCHER, M. J. & WEBER, A. P. 2013. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science,* 339**,** 1207-10.

SERBUS, L. R., CASPER-LINDLEY, C., LANDMANN, F. & SULLIVAN, W. 2008. The genetics and cell biology of *Wolbachia*-host interactions. *Annu Rev Genet,* 42**,** 683-707.

SHAND, R. F. L., K.J. 2008. Archaeal Animicrobials: an Undiscovered Country. *In:* BLUM, P. (ed.) *Archaea: New Models for Prokaryotic Biology.* Caister Academic Press.

SHEN, J., CURTIS, C., TAVARE, S. & TOWER, J. 2009. A screen of apoptosis and senescence regulatory genes for life span effects when over-expressed in *Drosophila*. *Aging (Albany NY),* 1**,** 191-211.

SHEWRY, P. R. 2003. Tuber storage proteins. *Ann Bot,* 91**,** 755-69.

SHI, Y., LI, N., YAN, Y., WANG, H., LI, Y., LU, C. & SUN, J. 2012. Combined Antibacterial Activity of Phage Lytic Proteins Holin and Lysin from *Streptococcus suis* Bacteriophage SMP. *Curr Microbiol,* 65**,** 28-34.

SHIBAYAMA, Y. & DABBS, E. R. 2011. Phage as a source of antibacterial genes: Multiple inhibitory products encoded by *Rhodococcus* phage YF1. *Bacteriophage,* 1**,** 195-197.

SHIMODAIRA, H. & HASEGAWA, M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution,* 16**,** 1114-1116.

SHOEMAKER, D. D., DYER, K. A., AHRENS, M., MCABEE, K. & JAENIKE, J. 2004. Decreased diversity but increased substitution rate in host mtDNA as a consequence of *Wolbachia* endosymbiont infection. *Genetics,* 168**,** 2049-58.

SHOEMAKER, D. D., KATJU, V. & JAENIKE, J. 1999. *Wolbachia* and the evolution of reproductive isolation between *Drosophilla recens* and *Drosophila subquinaria*. *Evolution,* 53**,** 1157-1164.

SHUTT, T. E. & GRAY, M. W. 2006. Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends Genet,* 22**,** 90-5.

SILVERMAN, J. M., BRUNET, Y. R., CASCALES, E. & MOUGOUS, J. D. 2012. Structure and regulation of the type VI secretion system. *Annu Rev Microbiol,* 66**,** 453-72.

SIOZIOS, S., IOANNIDIS, P., KLASSON, L., ANDERSSON, S. G., BRAIG, H. R. & BOURTZIS, K. 2013. The diversity and evolution of *Wolbachia* ankyrin repeat domain genes. *PLoS One,* 8**,** e55390.

SITKIEWICZ, I., STOCKBAUER, K. E. & MUSSER, J. M. 2007. Secreted bacterial phospholipase A2 enzymes: better living through phospholipolysis. *Trends Microbiol,* 15**,** 63-9.

SOUTHALL, T. D., ELLIOTT, D. A. & BRAND, A. H. 2008. The GAL4 System: A Versatile Toolkit for Gene Expression in *Drosophila*. *CSH Protoc,* 2008**,** pdb top49.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30**,** 1312-3.

SUMMER, E. J., BERRY, J., TRAN, T. A., NIU, L., STRUCK, D. K. & YOUNG, R. 2007. Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts. *J Mol Biol,* 373**,** 1098-112.

SZEKERES, S., DAUTI, M., WILDE, C., MAZEL, D. & ROWE-MAGNUS, D. A. 2007. Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection. *Mol Microbiol,* 63**,** 1588-605.

TANAKA, K., FURUKAWA, S., NIKOH, N., SASAKI, T. & FUKATSU, T. 2009. Complete WO phage sequences reveal their dynamic evolutionary trajectories and putative functional elements required for integration into the *Wolbachia* genome. *Appl Environ Microbiol,* 75**,** 5676-86.

TAYLOR, M., MEDIANNIKOV, O., RAOULT, D. & GREUB, G. 2011. Endosymbiotic bacteria associated with nematodes, ticks and amoebae. *FEMS Immunol Med Microbiol*.

TAYLOR, M. J., HOERAUF, A. & BOCKARIE, M. 2010. Lymphatic filariasis and onchocerciasis. *Lancet,* 376**,** 1175-85.

THEURETZBACHER, U. 2012. Accelerating resistance, inadequate antibacterial drug pipelines and international responses. *Int J Antimicrob Agents,* 39**,** 295-9.

TOOMEY, M. E., PANARAM, K., FAST, E. M., BEATTY, C. & FRYDMAN, H. M. 2013. Evolutionarily conserved *Wolbachia*-encoded factors control pattern of stem-cell niche tropism in *Drosophila* ovaries and favor infection. *Proc Natl Acad Sci U S A,* 110**,** 10788-93.

TRAM, U. & SULLIVAN, W. 2002. Role of delayed nuclear envelope breakdown and mitosis in *Wolbachia*-induced cytoplasmic incompatibility. *Science,* 296**,** 1124-6.

UDAY, J. & PUTTARAJU, H. P. 2012. Comparative analysis of *Wolbachia* surface protein in *D. melanoagster*, *A. tabida* and *B. malayi*. *Bioinformation,* 8**,** 711-5.

UEHARA, T. & BERNHARDT, T. G. 2011. More than just lysins: peptidoglycan hydrolases tailor the cell wall. *Curr Opin Microbiol,* 14**,** 698-703.

VALLENET, D., ENGELEN, S., MORNICO, D., CRUVEILLER, S., FLEURY, L., LAJUS, A., ROUY, Z., ROCHE, D., SALVIGNOL, G., SCARPELLI, C. & MEDIGUE, C. 2009. MicroScope: a platform for microbial genome annotation and comparative genomics. *Database (Oxford),* 2009**,** bap021.

VAN HAM, R. C., KAMERBEEK, J., PALACIOS, C., RAUSELL, C., ABASCAL, F., BASTOLLA, U., FERNANDEZ, J. M., JIMENEZ, L., POSTIGO, M., SILVA, F. J., TAMAMES, J., VIGUERA, E., LATORRE, A., VALENCIA, A., MORAN, F. & MOYA, A. 2003. Reductive genome evolution in *Buchnera aphidicola*. *Proc Natl Acad Sci U S A,* 100**,** 581-6.

VAN WOLFEREN, M., AJON, M., DRIESSEN, A. J. & ALBERS, S. V. 2013. How hyperthermophiles adapt to change their lives: DNA exchange in extreme conditions. *Extremophiles,* 17**,** 545-63.

VAVRE, F., FLEURY, F., LEPETIT, D., FOUILLET, P. & BOULETREAU, M. 1999. Phylogenetic evidence for horizontal transmission of *Wolbachia* in host-parasitoid associations. *Mol Biol Evol,* 16**,** 1711-23.

VAVRE, F., MOUTON, L. & PANNEBAKKER, B. A. 2009. *Drosophila*-parasitoid communities as model systems for host-*Wolbachia* interactions. *Adv Parasitol,* 70**,** 299-331.

VOLLMER, J., SCHIEFER, A., SCHNEIDER, T., JULICHER, K., JOHNSTON, K. L., TAYLOR, M. J., SAHL, H. G., HOERAUF, A. & PFARR, K. 2013. Requirement of lipid II biosynthesis for cell division in cell wall-less *Wolbachia*, endobacteria of arthropods and filarial nematodes. *Int J Med Microbiol,* 303**,** 140-9.

VOLLMER, W., JORIS, B., CHARLIER, P. & FOSTER, S. 2008. Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol Rev,* 32**,** 259-86.

VOTH, D. E., BROEDERDORF, L. J. & GRAHAM, J. G. 2012. Bacterial Type IV secretion systems: versatile virulence machines. *Future Microbiol,* 7**,** 241-57.

WALKER, T., JOHNSON, P. H., MOREIRA, L. A., ITURBE-ORMAETXE, I., FRENTIU, F. D., MCMENIMAN, C. J., LEONG, Y. S., DONG, Y., AXFORD, J., KRIESNER, P., LLOYD, A. L., RITCHIE, S. A., O'NEILL, S. L. & HOFFMANN, A. A. 2011. The wMel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. *Nature,* 476**,** 450-3.

WALKER, T., SONG, S. & SINKINS, S. P. 2009. *Wolbachia* in the *Culex pipiens* group mosquitoes: introgression and superinfection. *J Hered,* 100**,** 192-6.

WANG, I. N., SMITH, D. L. & YOUNG, R. 2000. Holins: the protein clocks of bacteriophage infections. *Annu Rev Microbiol,* 54**,** 799-825.

WANG, Z., SU, X. M., WEN, J., JIANG, L. Y. & QIAO, G. X. 2014. Widespread infection and diverse infection patterns of *Wolbachia* in Chinese aphids. *Insect Sci,* 21**,** 313-25.

WERREN, J. H., BALDO, L. & CLARK, M. E. 2008. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol,* 6**,** 741-51.

WERREN, J. H. & JAENIKE, J. 1995. *Wolbachia* and cytoplasmic incompatibility in mycophagous Drosophila and their relatives. *Heredity (Edinb),* 75 (Pt 3)**,** 320-6.

WERREN, J. H., ZHANG, W. & GUO, L. R. 1995. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proc Biol Sci,* 261**,** 55-63.

WHITE, R., CHIBA, S., PANG, T., DEWEY, J. S., SAVVA, C. G., HOLZENBURG, A., POGLIANO, K. & YOUNG, R. 2011. Holin triggering in real time. *Proc Natl Acad Sci U S A,* 108**,** 798-803.

WIEDENBECK, J. & COHAN, F. M. 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol Rev,* 35**,** 957-76.

WIMMER, P., SCHREINER, S. & DOBNER, T. 2012. Human pathogens and the host cell SUMOylation system. *J Virol,* 86**,** 642-54.

WOOLFIT, M., ITURBE-ORMAETXE, I., BROWNLIE, J. C., WALKER, T., RIEGLER, M., SELEZNEV, A., POPOVICI, J., RANCES, E., WEE, B. A., PAVLIDES, J., SULLIVAN, M. J., BEATSON, S. A., LANE, A., SIDHU, M., MCMENIMAN, C. J., MCGRAW, E. A. & O'NEILL, S. L. 2013. Genomic Evolution of the Pathogenic *Wolbachia* Strain, wMelPop. *Genome Biol Evol,* 5**,** 2189-204.

WU, B., NOVELLI, J., JIANG, D., DAILEY, H. A., LANDMANN, F., FORD, L., TAYLOR, M. J., CARLOW, C. K., KUMAR, S., FOSTER, J. M. & SLATKO, B. E. 2013. Interdomain lateral gene transfer of an essential ferrochelatase gene in human parasitic nematodes. *Proc Natl Acad Sci U S A,* 110**,** 7748-53.

WU, M., SUN, L. V., VAMATHEVAN, J., RIEGLER, M., DEBOY, R., BROWNLIE, J. C., MCGRAW, E. A., MARTIN, W., ESSER, C., AHMADINEJAD, N., WIEGAND, C., MADUPU, R., BEANAN, M. J., BRINKAC, L. M., DAUGHERTY, S. C., DURKIN, A. S., KOLONAY, J. F., NELSON, W. C., MOHAMOUD, Y., LEE, P., BERRY, K., YOUNG, M. B., UTTERBACK, T., WEIDMAN, J., NIERMAN, W. C., PAULSEN, I. T., NELSON, K. E., TETTELIN, H., O'NEILL, S. L. & EISEN, J. A. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol,* 2**,** E69.

XI, Z., GAVOTTE, L., XIE, Y. & DOBSON, S. L. 2008a. Genome-wide analysis of the interaction between the endosymbiotic bacterium *Wolbachia* and its *Drosophila* host. *BMC Genomics,* 9**,** 1.

XI, Z., KHOO, C. C. & DOBSON, S. L. 2005. *Wolbachia* establishment and invasion in an *Aedes aegypti* laboratory population. *Science,* 310**,** 326-8.

XI, Z., RAMIREZ, J. L. & DIMOPOULOS, G. 2008b. The *Aedes aegypti* toll pathway controls dengue virus infection. *PLoS Pathog,* 4**,** e1000098.

YAMADA, R., FLOATE, K. D., RIEGLER, M. & O'NEILL, S. L. 2007. Male development time influences the strength of *Wolbachia*-induced cytoplasmic incompatibility expression in *Drosophila melanogaster*. *Genetics,* 177**,** 801-8.

YAMADA, R., ITURBE-ORMAETXE, I., BROWNLIE, J. C. & O'NEILL, S. L. 2011. Functional test of the influence of *Wolbachia* genes on cytoplasmic incompatibility expression in *Drosophila melanogaster*. *Insect Mol Biol,* 20**,** 75-85.

YAMAGUCHI, Y., PARK, J. H. & INOUYE, M. 2011. Toxin-antitoxin systems in bacteria and archaea. *Annu Rev Genet,* 45**,** 61-79.

YE, Y. H., WOOLFIT, M., HUTTLEY, G. A., RANCES, E., CARAGATA, E. P., POPOVICI, J., O'NEILL, S. L. & MCGRAW, E. A. 2013. Infection with a Virulent Strain of Disrupts Genome Wide-Patterns of Cytosine Methylation in the Mosquito. *PLoS One,* 8**,** e66482.

YEN, J. H. & BARR, A. R. 1971. New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens L*. *Nature,* 232**,** 657-8.

YOUNG, I., WANG, I. & ROOF, W. D. 2000. Phages will out: strategies of host cell lysis. *Trends Microbiol,* 8**,** 120-8.

YOUNG, R. 2013. Phage lysis: do we have the hole story yet? *Curr Opin Microbiol,* 16**,** 790-7.

ZABALOU, S., APOSTOLAKI, A., PATTAS, S., VENETI, Z., PARASKEVOPOULOS, C., LIVADARAS, I., MARKAKIS, G., BRISSAC, T., MERCOT, H. & BOURTZIS, K. 2008. Multiple rescue factors within a *Wolbachia* strain. *Genetics,* 178**,** 2145-60.

ZELE, F., NICOT, A., BERTHOMIEU, A., WEILL, M., DURON, O. & RIVERO, A. 2014. *Wolbachia* increases susceptibility to *Plasmodium* infection in a natural system. *Proc Biol Sci,* 281**,** 20132837.

ZHANG, G., HUSSAIN, M., O'NEILL, S. L. & ASGARI, S. 2013. *Wolbachia* uses a host microRNA to regulate transcripts of a methyltransferase, contributing to dengue virus inhibition in *Aedes aegypti*. *Proc Natl Acad Sci U S A,* 110**,** 10276-81.

ZHAXYBAYEVA, O. & DOOLITTLE, W. F. 2011. Lateral gene transfer. *Curr Biol,* 21**,** R242-6.

ZHENG, Y., REN, P. P., WANG, J. L. & WANG, Y. F. 2011. *Wolbachia*-induced cytoplasmic incompatibility is associated with decreased Hira expression in male *Drosophila*. *PLoS One,* 6**,** e19512.

ZHOU, W., ROUSSET, F. & O'NEIL, S. 1998. Phylogeny and PCR-based classification of *Wolbachia* strains using wsp gene sequences. *Proc Biol Sci,* 265**,** 509-15.

ZUG, R. & HAMMERSTEIN, P. 2012. Still a host of hosts for *wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS One,* 7**,** e38544.

APPENDIX A. *W*MEL GENES MATCHING CI CANDIDATE CRITERIA

Table A1: *w*Mel genes meeting criterion 1a: genes absent or divergent in *w*Au, but conserved in *w*Ri and *w*Sim. Microarray data obtained from (Ishmael et al., 2009).

Locus	Function
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 (Eukaryotic DUF812?)
WD0579	hypothetical protein (virulence associated?)
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	hypothetical protein
WD0632	SUMO protease
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 A2. *w*Mel genes matching criterion 1b: genes present in the core genome of CI-inducing strains, but absent in the pan-genome of *w*Bm.

Locus	Function
WD0035	ankyrin repeat-containing protein
WD0038	Protein tolB
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
WD0255	transcriptional regulator, putative
WD0274	hypothetical protein
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
WD0286	ankyrin repeat-containing prophage LambdaW1
WD0288	prophage LambdaW1, site-specific recombinase resolvase family protein
WD0315	hypothetical protein
WD0324	hypothetical protein
WD0338	hypothetical protein
WD0345	RND family efflux transporter MFP subunit
WD0382	hypothetical protein

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
WD0507	DNA repair protein RadC, truncation
WD0508	transcriptional regulator, putative
WD0515	reverse transcriptase, interruption-C
WD0604	hypothetical protein
WD0623	transcriptional regulator, putative
WD0625	DNA repair protein RadC, putative
WD0626	transcriptional regulator, putative
WD0628	hypothetical protein
WD0631	hypothetical protein
WD0632	hypothetical 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
WD0645	reverse transcriptase, truncation
WD0686	hypothetical protein
WD0693	reverse transcriptase, putative
WD0696	hypothetical protein
WD0702	hypothetical protein
WD0713	hypothetical protein
WD0721	Mg chelatase-related protein
WD0724	hypothetical protein
WD0730	phosphatidylglycerophosphatase A, putative
WD0733	hypothetical protein

WD0748	hypothetical protein
WD0750	PQQ repeat-containing protein
WD0764	hypothetical protein
WD0787	araM protein
WD0790	hypothetical protein
WD0818	hypothetical protein
WD0823	hypothetical protein
WD0826	hypothetical protein
WD0835	hypothetical protein
WD0875	IS5 family transposase
WD0880	coenzyme PQQ synthesis protein C, putative
WD0883	dihydropteroate synthase, putative
WD0884	dihydrofolate reductase
WD0887	DNA repair protein RadA
WD0914	hypothetical protein
WD0932	IS5 family transposase
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
WD1118	hypothetical protein
WD1126	hypothetical protein
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 A3. *w*Mel genes matching criteria 2a: present in host gonads at the protein level.

Locus	Function	Notes
WD0001	chromosomal replication initiation factor dnaA	a
WD0065	HU-family DNA-binding protein	b
WD0278	minor tail protein Z	a
WD0421	tRNA-i(6)A37 methylthiotransferase miaB	a
WD0609	regulatory protein repA	a
WD0631	hypothetical protein	c
WD1063	wsp	a, b

a. *w*Mel homologs shown for *w*Ri MS hits on infected *D. simulans* sperm proteome, courtesy Daniel LePage.
b. *w*Mel homologs shown for *w*Pip MS hits on infected *Culex pipiens* mosquito gonads, from (Beckmann et al., 2013).
c. *w*Mel homologs shown for *w*Pip MS hits on infected *Culex pipiens* fertilized spermathecae and ovaries, from (Beckmann and Fallon, 2013).

Table A4. *w*Mel genes matching criterion 2b: present in host gonads at the RNA level. *w*Mel homologs of *w*VitA genes in *Nasonia vitripennis* ovaries transcriptome with >5 reads present, courtesy Lisa Funkhouser-Jones.

Locus	Function
WD0016	translation elongation factor G
WD0024	DNA-directed RNA polymerase
WD0039	metallo-beta-lactamase superfamily protein
WD0041	putative membrane protein
WD0147	tetratricopeptide repeat family protein
WD0292	ankryin repeat protein
WD0307	chaperonin GroL
WD0337	hypothetical protein
WD0550	ankryin repeat protein
WD0631	hypothetical protein
WD0632	hypothetical protein
WD0722	hypothetical protein
WD0745	putative outer membrane protein
WD0838	hypothetical protein
WD0906	S1 RNA binding domain protein
WD0928	chaperone protein DnaK
WD0950	uncharacterised protein family UPF0005
WD1063	outer surface protein
WD1064	RNA polymerase sigma-32 factor
WD1071	cytochrome b
WD1202	DNA gyrase
WD1236	DNA/RNA helicase
WD1238	fructose-bisphosphate aldolase class 1
WD1249	sodium/hydrogen exchanger family protein
WD1278	hypothetical protein

Table A5. *w*Mel genes matching criterion 3: predicted to be secreted by the type IV secretion system.

Locus	Function
WD0026	hypothetical protein
WD0035	ankyrin repeat domain protein
WD0073	ankyrin repeat domain protein
WD0147	ankyrin repeat domain protein
WD0285	prophage LambdaW1, ankyrin repeat domain protein
WD0286	prophage LambdaW1, ankyrin repeat domain protein
WD0291	prophage LambdaW1, ankyrin repeat domain protein
WD0292	prophage LambdaW1, ankyrin repeat domain protein
WD0294	ankyrin repeat domain protein
WD0385	ankyrin repeat domain protein
WD0424	hypothetical protein
WD0430	hypothetical protein
WD0438	ankyrin repeat domain protein
WD0441	ankyrin repeat domain protein
WD0462	hypothetical protein
WD0498	ankyrin repeat domain protein
WD0514	ankyrin repeat domain protein
WD0549	translocase
WD0550	ankyrin repeat domain protein
WD0570	prophage P2W3, tail tape measure protein
WD0615	hypothetical protein
WD0633	prophage LambdaW5, ankyrin repeat domain protein
WD0636	prophage LambdaW5, ankyrin repeat domain protein
WD0637	prophage LambdaW5, ankyrin repeat domain protein
WD0672	ribosomal protein S17
WD0696	hypothetical protein
WD0706	hypothetical protein
WD0754	ankyrin repeat domain protein
WD1237	ATP-dependent Clp protease, ATP-binding subunit ClpA
WD1314	Fic family protein

Table A6. *w*Mel genes matching criterion 4: previously tested in transgenic *D. melanogaster*. Experimental tests of involvement in CI are described in (Yamada et al., 2011).

Locus Tag	Function
WD0294	Ankyrin repeat domain protein
WD0385	Ankyrin repeat domain protein
WD0498	Ankyrin repeat domain protein
WD0514	Ankyrin repeat domain protein
WD0550	Ankyrin repeat domain protein
WD0579	Gp16 similar to VrlC
WD0580	Gp15 similar to VrlC
WD0594	Prophage λW4, DNA methylase
WD0633	Prophage λW5, ankyrin repeat domain protein
WD0636	pk2, prophage λW5, ankyrin repeat domain protein
WD0754	Ankyrin repeat domain protein
WD0776	Ankyrin repeat domain protein