Molecular Mechanism of *Wolbachia* Induced Cytoplasmic Incompatibility

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

John Frederick Beckmann

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Ann Fallon, Advisor

May 2014

© John Frederick Beckmann 2014

Acknowledgements

First and foremost I would like to thank Dr. Ann Fallon, my advisor, for seeing scientific potential in me and for granting me the opportunity to pursue the research contained herein. Dr. Fallon has been extremely generous to me, not only allowing me to begin research under her tutelage despite my lackluster GRE/MCAT scores, unorthodox entry into the graduate program, and stubborn independence; but furthermore, has allowed me unlimited freedom to pursue my own passions, ideas, and experiments, many of which may have been too risky for non-naïve individuals. She has allowed me to work on my own terms and schedule, including sporadically coming and leaving as I please, often times in sweatpants and t-shirts with stains and holes. She has supported my children and family financially as well as mentally, allowing me to bring my children to work and lab meetings. Scientifically, Dr. Fallon has one of the sharpest minds in the department and has contributed countless ideas and thoughts which have helped to hone my research skills and hypotheses. Her expertise and historical knowledge have been priceless. Without Dr. Fallon's limitless support none of this research would have happened.

My committee members Dr. Tim Kurtti, Dr. Gerald Baldridge, and Dr. Stephen Katz have also been extremely generous and made themselves available for mentorship and review. I first met Dr. Katz in my undergraduate physiology class. There, he was notably one of the more memorable and energetic teachers who made kidney biology fun by entertaining somewhat inappropriate urination questions from me. Later when I was applying for medical school and physiology graduate school, Dr. Katz provided positive and insightful support that had a significant effect on my career and displayed his kind nature. He talked to me about my research on mosquitoes (which I knew very little about at the time), told me that he remembered me from his class, and offered me career advice, stating that the best way to get into medical school or a graduate program was to continue to "Research, Research, Research." I am now very grateful for receiving this advice because it motivated me to continue pursuing that path, which at the time seemed uncertain. Dr. Katz left an impression on me as being extremely supportive, despite me not being accepted into the physiology program or medical school. I am very grateful that Dr. Katz later served on my committee.

When I joined the Department of Entomology Dr. Kurtti was the first person to introduce himself to me and has since been a very accessible, honest, and useful mentor both here and away at scientific meetings. Dr. Kurtti has allowed me unlimited access to his supplies, instruments, and personnel. Dr. Kurtti and Dr. Uli Munderloh lead a journal club which has expanded my knowledge of closely related fields and probed the strength of my hypotheses and criticisms.

Dr. Baldridge has been both a mentor and personal friend who strongly contributed to my success. He has taught me techniques, reviewed my writings in depth, and offered ideas of his own as well as provided me with invaluable personal and career advice. Above all Dr. Baldridge inspires great confidence in me by always looking past my ignorance and entertaining the idea that I am an intellectual equal worthy of discussions related to technique, data, and *Rickettsiology* in general.

Aside from my committee I am extremely thankful for the researchers and co-workers who have made their time available to me including but not limited to: Dr. Mark Hochstrasser, who has offered his support and agreed to let me continue this research at Yale. Rod Felsheim, who I have pestered incessantly, sought advice from, run ideas by, and forced to teach me bioinformatics. Rod has been invaluable as the person who I saw as most knowledgeable in the techniques I was interested in. Anna Gerenday, who has listened to my complaining and provided positive motivation and a kick in the butt at times when I was very upset. Todd Markowski, LeeAnn Higgins, and Bruce Witthuhn at the University of Minnesota's Center for Mass Spectrometry and Proteomics who have not only run numerous samples of mine, but have always gone out of their way to brainstorm, answer questions, and plan experiments. The students in our lab who have allowed me the opportunity to mentor, teach, and learn from them including Pang Xiong Elissa Carroll, and Grace Li. Cassandra Kurtz for taking care of the mosquitoes and lab. My friends and fellow graduate students Corey Mcqueen, Kit Martin, and Geoff Lynn for providing support, motivation, and enjoyable lunches, dinners, and parties filled with drinks, gossip, gambling, and complaining.

Last but not least I would like to thank God and my family. My parents, David Beckmann and Sarah Beckmann, have always provided me with unlimited homebrewed beer (post-age 21), love, support, and education, without which this work would have been impossible. They always made it clear to me that I was to live up to their high expectations. My sisters inspire me to compete against their numerous successes. My grandmother, Gerry Beckmann, has taken me out to eat countless times providing encouragement and advice. My daughters, Jane, Elaine, and future children motivate me and make me happy every day. My wife and best friend, Shawna Beckmann, has provided incredible mental and spiritual support throughout this process. She has cooked our meals, fed our dog, changed the oil in our cars, filed our taxes, cleaned our house, bore our children, and believed in my delusions of grandeur all while quitting her job, starting and expanding her own business; she yearly makes far more money than me and is a true feminist in the most real sense!

This research was supported by NIH grant AI081322 and by the University of Minnesota Agricultural Experiment Station, St. Paul, MN.

Dedication

This thesis is dedicated to Dr. Hannes Laven who was the first to foresee the potential of cytoplasmic incompatibility, pioneered its biological and cytological study, and proved its applicable use by eradicating an elephantiasis vector in Burma; more impressively all of these studies were published in the same year, 1967.

Abstract

Wolbachia are obligate intracellular endosymbionts which live in the gonads of many arthropods of economic and medical importance. In insects, *Wolbachia* manipulate reproduction in a way that favors the spread of their infection. Cytoplasmic Incompatibility (CI), is a particular effect induced by *Wolbachia* infection in mosquitoes and other insects. CI causes conditional male sterility and produces a selective pressure in mixed populations of infected and uninfected mosquitoes giving *Wolbachia*-infected females a reproductive advantage. CI has been proposed as a gene drive tool which could be used to replace wild arthropod disease vectors with genetically modified ones less capable of transmitting diseases. CI has been demonstrated to be an effective agent at manipulating vector populations in the wild. When I began my research on *Wolbachia* in 2009, a central unresolved question, which has remained unanswered since the 1950's, concerned the molecular basis of CI; my doctoral research has wholly focused on answering this basic question, "What is the *Wolbachia* gene/protein that induces CI in mosquitoes?"

Table of Contents

List of Tables……pg. vii

List of Figures……..pg. viii

Chapter 1. A History of *Wolbachia* and Cytoplasmic Incompatibility……...…pg. 1
Initial identification and classifications of *Rickettsia* and *Wolbachia*……...pg. 1
Origins of Research on Cytoplasmic Incompatibility……....pg. 4
CI Population Dynamics……pg. 7
Cytological Studies of CI……pg. 11

Chapter 2. Decapitation Improved Detection of *Wolbachia Pipientis* (*Rickettsiales: Anaplasmataceae*) in *Culex pipiens* Linnaeus (*Diptera: Culicidae*) Mosquitoes by the Polymerase Chain Reaction……..pg. 16
Preface to Chapter 2……pg. 16
Summary……pg. 16
Introduction……..pg. 18
Materials and Methods……pg. 19
Results……..pg. 22
Discussion……pg. 35

Chapter 3. Detection of the Wolbachia-Encoded DNA Binding Protein, HU beta, in Mosquito Gonads……...pg. 39
Preface to Chapter 3……pg. 39
Summary……pg. 40
Introduction……pg. 41
Materials and Methods……pg. 43
Results……pg. 49
Discussion……pg. 67

Chapter 4. Detection of the *Wolbachia* Protein WPIP0282 in Mosquito Spermathecae: Implications for Cytoplasmic Incompatibility……..……pg. 74
Preface to Chapter 4……pg. 74
Summary..……....pg. 75
Introduction……..pg. 76
Materials and Methods……pg. 79
Results……..pg. 83
Discussion……..pg. 105
Additional Unpublished Proteomic Results……...pg. 113

Chapter 5. Cloning and Characterization of the *Wolbachia* Spermatheca Associated Protein WPIP0282 in a Highly Repressible Expression System for Potentially Toxic Proteins……pg. 119

Introduction……...pg. 119
Materials and Methods……..pg. 121
Results……pg. 126
Discussion…….....pg. 141
Chapter 6. Final Hypotheses and In-Silico Testing of the WPIP0282/0283 Toxin-Antidote Theory of Cytoplasmic Incompatibility……pg. 143
Introduction……pg. 143
Results and Discussion……..pg. 145

Bibliography……pg. 159

List of Tables

Chapter 3.
Table 1……pg. 54

Chapter 4.
Table 1……pg. 116

Chapter 5.
Table 1……pg. 124
Table 2……pg. 132

Chapter 6.
Table 1……pg. 150

List of Figures

Chapter 1.
Fig 1……..pg. 8

Chapter 2.
Fig 1……..pg. 24
Fig 2……..pg. 28
Fig 3……..pg. 30
Fig 4……..pg. 32

Chapter 3.
Fig 1……..pg. 50
Fig 2……..pg. 52
Fig 3……..pg. 58
Fig 4……..pg. 60
Fig 5……..pg. 64
Fig 6……..pg. 68

Chapter 4.
Fig 1……..pg. 86
Fig 2……..pg. 88
Fig 3……..pg. 92
Fig 4……..pg. 96
Fig 5……..pg. 98
Fig 6……..pg. 100
Fig 7……..pg. 106

Chapter 5.
Fig 1……..pg. 130
Fig 2……..pg. 136
Fig 3……..pg. 138

Chapter 6.
Fig 1……..pg. 148
Fig 2……..pg. 156

CHAPTER 1:

A HISTORY of *WOLBACHIA* and CYTOPLASMIC INCOMPATIBILITY

Initial Identification and Classifications of Rickettsia and Wolbachia.

There were two independent areas of scientific inquiry which gave rise to the initial and earliest work on the *Wolbachia* bacterium. These two areas of interest were research in the early 1900's on Rocky Mountain Spotted Fever (RMSF), a bacterial disease caused by infection with *Rickettsia rickettsii,* and research on the mysterious phenomenon of cytoplasmic incompatibility, a reproductive manipulation causing sterility in insects, which revitalized interest in *Wolbachia* in the 1960's.

Scientific work on RMSF was initiated by Louis B. Wilson and William M. Chowning in the Bitterroot Valley of Montana in 1902. These researchers were drawn to the allure of identifying and characterizing an unknown disease plaguing the western settlers. The disease itself was very troublesome and famously causes fever, maculopapular rashes, petechial rashes, and before antibiotics showed a very high mortality rate of 72.73%. Mysteriously, the disease seemed only to occur on the western side of the Bitterroot river. Initially Wilson and Chowning performed an epidemiological survey revealing most cases did in fact occur on the west side of the river. Wilson and Chowning then began performing microscopic studies in the endemic areas which led them to immediately suspect that ticks were the vectoring the disease. However, the two

did not correctly identify the causative bacteria and so the tick hypothesis was still a point of debate many years after (Harden, 1990).

In 1906 Howard Taylor Ricketts independently began working on RMSF and had a breakthrough when he discovered that the pathogen could be extracted from infected patients and cultured within guinea pigs. During the years 1907-1909 Ricketts verified tick-borne transmission of the disease by demonstrating transmission from one guinea pig to another via the Rocky Mountain Wood Tick, *Dermacenter andersoni.* Ricketts also produced a successful antiserum and observed the pathogen directly with Giemsa stain. Unfortunately, Ricketts' career was cut short when he became infected studying typhus in Mexico. He died in 1910 (Harden, 1990).

Ricketts' work spawned derivative studies from scientific followers who sought to identify and characterize the organisms that lived in arthropods. Marshall Hertig and Simeon Burt Wolbach continued Ricketts' work and identified the pathogen of RMSF. Wolbach proposed the name *Rickettsia rickettsii* in honor of the deceased Ricketts and described the bacteria as "a new type of parasite" in 1919 (Harden, 1990). However, Rocha-Lima in 1916 was actually the first to coin the term *Rickettsia*, again in honor of Ricketts, but in this case with respect to the causative agent of typhus, which lived in lice. In this period the definition of the term *Rickettsia* simply meant "minute bacillary forms" found in arthropods, some of which caused significant disease (Rocha-Lima, 1916).

Because the definition of "*Rickettsia"* was so ambiguous, the main goal of researchers at this point was to further identify as many Rickettsial organisms in as many arthropods as possible with the broader goal of determining what it really meant to be

classified as a *Rickettsia*. In 1922, Cowdry noted that *Rickettsia* organisms were mostly contained within arthropods that fed on blood as compared to other diets. There was also debate about whether or not *Rickettsia* lived intracellularly (Cowdry, 1923). In 1924 Hertig and Wolbach showed that *Rickettsia* were indeed intracellular organisms which lived and were maternally transmitted in the reproductive tissues of arthropods. However, *Rickettsia*, they noted, were somehow distinct from other intracellular organisms such as the "bacteroids of cockroaches, and the yeasts of aphids and other hemiptera." The studies in 1924 by Hertig and Wolbach were the first to visually identify *Wolbachia* organisms, however at this time they were simply described as Gram negative rod-like coccoid bodies found in the smears of ovaries and testes in *Culex pipiens* mosquitoes and classified as harmless *Rickettsia* (Hertig and Wolbach, 1924). The results of these surveys led the field to further define *Rickettsia* as "Gram-negative, intracellular, bacterium-like organisms found in arthropods which are very small in diameter, less than 0.5 microns, coccoid, and difficult to culture (Cowdry, 1923; Rocha-Lima, 1916).

In 1936 a subsequent study by Hertig (himself a graduate of the University of Minnesota's Department of Entomology) led to the first true definition of a *Wolbachia* organism. Hertig defined *Wolbachia* as something different from *Rickettsia* which had "peculiar morphologies," or more specifically, they were pleomorphic. Hertig then proposed the name *Wolbachia pipientis* in honor of Wolbach. At this point *Wolbachia* represented the harmless *Rickettsia*-like organisms found in *Culex pipiens* which were morphologically different and yet classified within the broader category of *Rickettisa*. Hertig was also the first to note that *Wolbachia* infect testes, but are removed from sperm

(Hertig, 1936), an interesting fact that would become important for later *Wolbachia* research. At this point research on *Wolbachia* by scientists interested in *Rickettsia* ceased; presumably because they were characterized as "harmless" and since they were somehow different from *Rickettsia* were no longer of interest to Rickettsiologists. However, work on *Wolbachia* would unknowingly later pick up due to entomologists interested in the biological phenomenon of cytoplasmic incompatibility in mosquitoes.

Origins of Research on Cytoplasmic Incompatibility.

Cytoplasmic incompatibility, CI, is a conditional sterility that occurs within insects and is induced by *Wolbachia*. The first documentation of CI occurred in 1938 by Marshall who noted that British *Culex pipiens* mosquitoes would not produce successful offspring when mated to *Culex pipiens* mosquitoes from France. Marshall noted that eggs were laid and yet were somehow prevented from hatching (Marshall, 1938). However, research on CI did not really begin until the 1950's When Hannes Laven began to focus his work on its study for the purposes of understanding arthropod speciation and possibly using it as a tool for the biological control of insects. Laven's first descriptions of CI were in papers published in German in 1953 and 1957 (Laven, 1953; 1957). In 1967 Laven published an English paper describing a new landmark concept of evolution whereby speciation doesn't necessarily occur only by geographical reproductive isolation and buildup of genetic differences (the belief at the time), but could also be induced by cytoplasmic factors that limit reproduction, citing CI as a prime example (Laven, 1967a). He also claimed that in fact, this force may be more important, effective, and faster than

spatial isolation because single mutations in some unknown cytoplasmic factor would produce complete reproductive isolation of populations.

A subsequent paper by Laven in 1967 sought to define and identify these evolutionary cytoplasmic factors by performing numerous mosquito crossing experiments from strains of *Culex pipiens* isolated in 39 different regions of Europe. Many of these crosses would not produce viable offspring and Laven himself was astounded that crosses from *Culex pipiens* in Hamburg, Germany, were incompatible with *Culex pipiens* in Paris, France. This study revealed what Laven described as the 17 different cytotypes of *Culex pipiens*, some of which were bi or uni-directionally incompatible (Laven, 1967b). Laven further biologically characterized the phenomenon by demonstrating that the sperm entered the eggs successfully and therefore was not limited by a physical barrier. However, karyogamy (the process of the paternal and maternal nuclei combining) was prevented and the paternal genome did not contribute in the development of the inviable embryos. Because karyogamy was prevented, Laven ruled out mosquito chromosomal genes as a factor and determined that the factor must be cytoplasmic, thus coining the term "cytoplasmic incompatibility." Laven was the first to argue that the cytoplasmic factor was likely an independent genome of viral or bacterial origins (Laven, 1967b).

A third landmark paper published again in 1967 by Laven demonstrated that the sterility induced by cytoplasmic incompatibility could be harnessed for biological control (Laven, 1967c). In this study, Laven eradicated a vector of filariasis, *Culex pipiens fatigans,* in Burma (now Myanmar) by releasing vast numbers of incompatible males. This technique is a form of the sterile insect technique (More typically, SIT involves use

of irradiated males). This study was the major catalyst for most subsequent work on *Wolbachia*, which was justified by using the (then unknown) cytoplasmic factor as a means of arthropod population manipulation. Interestingly, it was not until 1971, that Yen and Barr proposed the hypothesis that the bacteria, discovered by Hertig and Wolbach in 1924, was the causative agent of cytoplasmic incompatibility (Yen and Barr, 1971). Later in 1973 this same group proved this hypothesis by correlating CI with antibiotic treatment and the removal of *Wolbachia*; the absence of *Wolbachia* in males causes them to become compatible. They also characterized the rescue process of CI and correlated this with *Wolbachia* presence (Yen and Barr, 1973). In summary, what was known at this point was that *Wolbachia* live within the testes and ovaries of arthropods and express some cytoplasmic factor that causes sterility when infected males mate with uninfected females. However, if the infected male mated with an infected female, there was a secondary cytoplasmic factor that was able to correct or fix whatever was originally altered; in this way CI was known to be like a toxin-antidote system. It was understood that this toxin-antidote based sterility should produce an evolutionary advantage favoring the spread of the cytoplasmic factor and/or *Wolbachia* because females infected with *Wolbachia* are compatible with both infected and non-infected males whereas uninfected females are only compatible with uninfected males. The cytological relationships of CI are described in Figure 1.

CI Population Dynamics.

Once the driving cytoplasmic factor of CI (*Wolbachia*) had been found in multiple insects including *Drosophila*, which was coincidentally found to be infected with *Wolbachia* by serendipitous detection of its genome during *Drosophila* genome sequencing, research on CI expanded and split into two fields. The first field was composed of researchers who sought to understand and quantify the mathematical principles behind its evolutionary drive and the second was composed of researchers who sought to understand the cytological mechanics of CI.

During the period in which Laven was arguing that CI was a new means of evolutionary speciation an opposing hypothesis was proposed. In 1959, Caspari and Watson argued that cytoplasmic incompatibility should not induce speciation events because the cytoplasmic factor should itself spread throughout the population until it reached fixation and there would no longer be two distinct populations but rather one population with a universal cytoplasmic factor. They published a mathematical model describing the spread and fixation of such a cytoplasmic factor: one that would induce complete sterility between competing populations. In effect this model was somewhat proven correct in Laven's 1967 replacement of *Culex pipiens fatigans* in Burma, described above. However, the mathematical model itself turned out to be slightly too simplistic and was later modified in 1978 by Paul Fine to include maternal transmission and fitness advantages/disadvantages within the insects. This new revised model led Fine to propose the novel idea that the genetic drive induced by CI might be powerful enough to drive deleterious genes/genotypes to fixation in a population. In essence, this was the

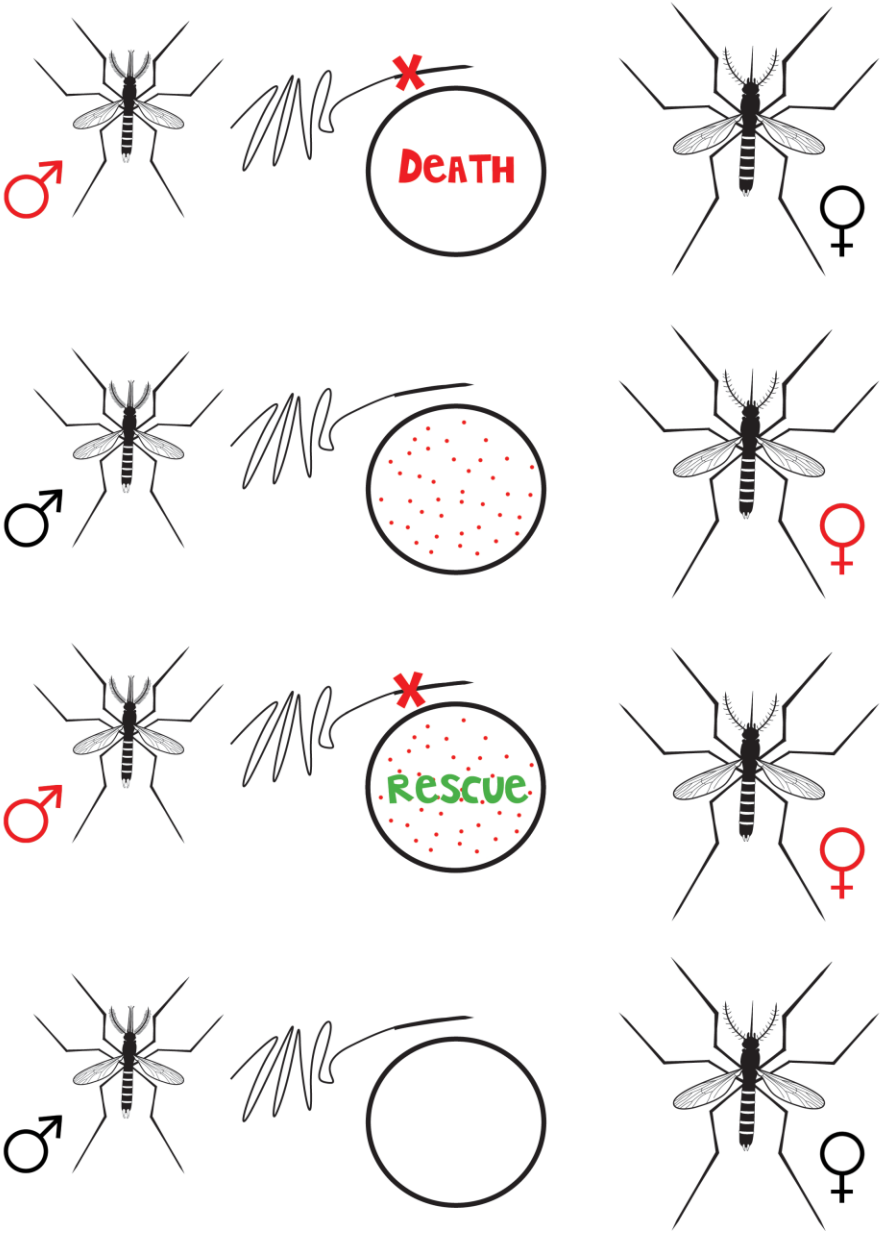

DEATH
RESCUE

Figure 1. Cytological relationships of *Wolbachia* infection and mating. Red male and female symbols represent infected mosquitoes. Each infected male produces sperm that is modified by *Wolbachia.* Females infected with *Wolbachia* produce eggs infected with *Wolbachia*. *Wolbachia* infected eggs express a rescue factor capable of rescuing *Wolbachia* modified sperm and are also viable when fertilized with normal sperm. CI induces embryonic death when modified sperm fertilize uninfected female eggs that do not have the rescue factor. A driving genetic force is produced because infected females have an advantage over uninfected females in that they can produce viable offspring regardless of the infection status of their mate.

first proposal of using *Wolbachia* as a means to drive disease resistant transgenes, which themselves can be deleterious to the organism's fitness, in vector populations. At this point *Wolbachia* was no longer just an alternative means of reducing vector populations, but now was proposed to be a way to specifically manipulate or to pick and choose which populations of insects survived and thrived in the wild.

One discussion within the *Wolbachia* population dynamics community at the time was whether or not the genetic drive of CI was truly powerful enough to replace uninfected populations. Evidence seemed to suggest that CI was in fact a weak driver because incompatibilities seemed to persistently remain within populations (Barr, 1980). The fact that incompatibilities persisted, seemed to disprove the Caspari and Watson model, which predicted that *Wolbachia* infection should reach fixation, at which point there would no longer be incompatibilities. A stroke of luck then occurred in California which presented an opportunity to resolve these issues. In 1986 Hoffmann, Turelli, and Simmons happened upon discovering a strain of *Wolbachia* infected *Drosophila simulans* which seemed to be on the rise. The researchers then made the decision to follow this strain to see whether its *Wolbachia* infection would spread throughout California. If the *Wolbachia* infection spread to fixation it would prove that the genetic drive of CI was really happening in nature (Hoffmann et al., 1986). The group continued studies on the organism and in 1991 noted that the *Wolbachia* infected *Drosophila simulans* strain had indeed spread into northern California at a rate of about 100 km/year (Turelli and Hoffman, 1991). The studies by these researchers during this period demonstrated that *Wolbachia* induced CI does possess the necessary genetic drive to induce its spread.

However there was still debate within the community about the strength of this force and which factors exacerbate and weaken it. Recent studies have continued this debate and have shown that the power of the genetic drive of CI is dependent upon factors such as the strength of the incompatibility, hatch rates, maternal transmission, initial infection rates, and fecundity, all of which can vary with differing *Wolbachia* strains, different arthropod hosts, and in different field conditions (Hoffmann et al., 1990; Kriesner et al., 2013). However, it is generally accepted now that *Wolbachia* infections, once reaching a certain size will tend to spread. This phenomenon has been further demonstrated in other locations such as Australia (Kriesner et al., 2013).

Cytological Studies of CI.

One of the earliest significant cytological studies on CI was published in 1968 by Ryan and Saul even before *Wolbachia* had been shown to be the cytoplasmic factor of CI. This study was the first to specifically highlight biological phenotypes associated with CI in embryonic insect cells. Specifically, the researchers noted that incompatible embryos from the wasp *Nasonia vitripennis* showed paternal chromosomes forming as tangled masses rather than chromosomes. The researchers speculated that this was caused by some cytoplasmic factor inherited through the eggs. In 1970, Erich Jost repeated Laven's earlier work by using genetic markers to demonstrate that in CI embryos from *Culex pipiens* karyogamy is inhibited, preventing the paternal chromatin from contributing to embryonic development. These two studies were the first observations that embryonic

death from CI was somehow caused by or produced defects in the chromatin of insect sperm.

As was discussed above, Yen and Barr demonstrated that CI was correlated with the presence of *Wolbachia* infecting *Culex pipiens* in 1971-1973, after which, people immediately began noting that the bacterial density of *Wolbachia* infections seemed to also correlate with the strength of CI and embryonic death rates. Subarrao et al. (1977) was the first to propose this hypothesis as the density model of CI. Later, in 1987, Richardson et al. demonstrated that the CI in *Nasonia vitripennis* was similarly correlated with the presence and absence of *Wolbachia* organisms. At this point CI had been correlated with *Wolbachia* in *Culex* and *Nasonia*; people then began to assume that other cases of CI in other insects were also likely caused by *Wolbachia*. In this paper Saul also presented a sperm hypothesis stating that CI is likely to be caused by an "antigenic effect conferred upon the sperm."

From then on the 1990's saw a period of researchers trying to identify what *Wolbachia* did to insect sperm. In 1990 Breeuwer and Werren were the first researchers to report that the embryonic death resulting in CI crosses of *Nasonia* was a result of improper paternal chromatin condensation. Also in 1990 O'Neill and Karr reiterated this point stating that in *Drosophila simulans*, the cellular basis of CI involves disruption of processes before or during zygote formation and it arises from defects in the structure and function of the sperm.

The first study analyzing the cell cycles of CI embryos was published in 1995 by Reed and Werren. Here the researchers described paternal chromatin from CI embryos in

Nasonia vitripennis as less dense than the maternal chromatin; they then followed the fate of this chromatin as it progressed through multiple cell cycles noting that it formed a tangled mass and was sheared by subsequent divisions. This paper was the first to propose the timing model for CI which states that the paternal chromatin's cell cycle is simply delayed as compared to the maternal chromatin; subsequently, the egg, which follows signals from the maternal chromatin, advances through mitosis normally, while the paternal chromatin lags behind and is sheared at anaphase. The timing model of CI implicates that *Wolbachia* are able to rescue CI by simply slowing down the maternal chromatin's mitotic cell cycle or conversely by speeding up the paternal's. Essentially the timing model says that the paternal and maternal nuclei in the zygote are at different phases in their respective cell cycles. In 1997 independent experiments by Calliani et al. in *Drosophila simulans* reiterated that paternal chromosomes failed to condense properly and uniquely mentioned that they lagged behind on the metaphase plate while the maternal chromosomes themselves entered anaphase. This paper specifically strengthened the timing hypothesis of CI by observing that the paternal chromatin in fact continued to condense during anaphase and occasionally would eventually form chromosomes suggesting that some effect had simply slowed or paused the chromatin condensation process momentarily. Calliani et al. 1997 then hypothesized that *Wolbachia* release a factor which is associated with male chromatin during spermatogenesis and causes an effect on chromatin scaffolding proteins such as topoisomerases, which are themselves important for the formation of mitotic/meiotic chromosomes.

In 1999 Daven Presgraves published a unique study utilizing gynogenetic stocks of *Drosophila melanogaster* to demonstrate that *Wolbachia* unequivocally induce a modification only on the paternal chromatin of sperm as opposed to some other extranuclear sperm protein or signaling factor in the cytoplasm/nucleoplasm. Gynogenesis is the process by which a zygote can develop without genetic contribution from the males. In this way gynogenesis is similar to parthenogenesis. However, gynogenetic eggs do still require extranuclear factors from the sperm to develop. Therefore, Presgraves mated *Wolbachia* with modified sperm to gynogenetic females with the following logic, "If *Wolbachia* disrupt paternal chromosomes only, diploid gynogenetic eggs should develop; if, however, *Wolbachia* disrupt any extranuclear paternal factors required for development, diploid gynogenetic eggs should not develop." The results of the study were that the incompatible CI matings produced perfectly viable offspring proving that *Wolbachia* specifically induce some kind of epigenetic effect on sperm chromatin only.

In 2002-2003, Michael Clark and colleagues published a series of studies microscopically studying *Wolbachia* infections in *Drosophila* testes. Specifically the researchers followed the infections throughout sperm development and differentiation, again showing that *Wolbachia* are removed from adult sperm during the spermatid stage and positing the logical theory that whatever the epigenetic effect on sperm was, it had to be induced before be spermatid phase when *Wolbachia* are removed. After this phase the effect remained upon the sperm until fertilization.

Current hypotheses of CI are based on the above data. For the most part, there have not been any truly significant advances with respect to understanding how *Wolbachia* induce CI within the last 15 years. However, the field currently believes that CI results from a *Wolbachia*-mediated modification on developing spermatocyte/spermatid chromatin (Clark et al., 2003; Presgraves, 1999) that can be rescued when the male pronucleus matures in the cytoplasm of eggs from infected females. In eggs from uninfected females, sperm modification in the absence of the rescue factor disrupts cell cycle synchrony between male and female pro-nuclei (Serbus et al., 2008; Callaini et al., 1997); in diploid insects such as mosquitoes, eggs from a CI cross fail to hatch, due to improper condensation of the paternal chromatin and subsequent entanglement/severing during the first anaphase of the zygote. The subsequent work included within this thesis represents my efforts to elucidate the unknown epigenetic effector/CI inducer and how it mechanistically functions.

The discussion above was written to give a historical perspective and highlight significant advances within the timeline of *Wolbachia* research as well as provide a short biological context for the subsequent chapters. No doubt, I have not included many research publications whose authors would feel that they contributed to the field. However, I sought to only include contributions that were either the first to propose a particular idea or ones that were particularly insightful and well crafted. For more reading on the biological hypotheses of CI, insect fertilization, and insect development, an excellent review has already been written: Serbus et al. 2008, which includes all the most recent hypotheses excepting those within the subsequent thesis chapters.

CHAPTER 2:

DECAPITATION IMPROVED DETECTION OF *WOLBACHIA PIPIENTIS* (*RICKETTSIALES: ANAPLASMATACEAE*) IN *CULEX PIPIENS* LINNAEUS (*DIPTERA: CULICIDAE*) MOSQUITOES BY THE POLYMERASE CHAIN REACTION

***Preface to Chapter 2*.** The following is a reproduction of a publication from the Journal of Medical Entomology: Sep 2012; 49(5): 1103–1108 authored by John F. Beckmann and Ann M. Fallon. This research was initiated under the premise of developing and characterizing colonies of both *Wolbachia* infected and tetracycline-cured *Culex pipiens* mosquitoes for future proteomic and genetic comparisons. This publication documents the problems I encountered and solutions developed when screening these mosquitoes to verify infection by the polymerase chain reaction. This research was the first necessary step in a series of experiments undertaken to investigate the molecular basis of CI.

Summary:

Polymerase chain reaction (PCR) is often used to detect microorganisms, pathogens, or both, including the reproductive parasite *Wolbachia pipientis* (Rickettsiales: Anaplasmataceae), in mosquitoes. Natural populations of *Culex pipiens* L. (Diptera: Culicidae) mosquitoes are infected with one or more strains of *W. pipientis*, and crosses between mosquitoes harboring different *Wolbachia* strains provide one of the

best-known examples of cytoplasmic incompatibililty (CI). When we used PCR to monitor *Wolbachia* in the Buckeye strain of *Culex pipiens*, and in a *Wolbachia*-cured sister colony obtained by tetracycline treatment, we noted false negative PCR reactions with DNA samples from infected mosquitoes; these results were inconsistent with direct microscopic observation of *Wolbachia*-like particles in gonads dissected from mosquitoes in the same population. Assays with diluted template often improved detection of positive samples, suggesting that DNA prepared from whole mosquitoes contained an inhibitor of the PCR reaction. We reconciled discrepancies between PCR and microscopy by systematic measurement of the PCR reaction in the presence of an internal standard. Mosquito decapitation before DNA extraction restored the reliability of the PCR reaction, allowing accurate determination of *Wolbachia* infection status in infected and tetracycline-cured mosquito populations, consistent with microscopic examination. Using PCR primers based on the *Tr1* gene, we confirmed that the *Wolbachia* infection in the Buckeye strain of *Culex pipiens* belongs to the genotype designated *w*Pip1. Finally, to explore more widely the distribution of PCR inhibitors, we demonstrated that DNA isolated from the cricket, *Acheta domesticus* (L.); the beetle, *Tenebrio molitor* L.; the honey bee, *Apis mellifera* L.; and the mosquito, *Anopheles punctipennis* Say also contained PCR inhibitors. These results underscore the importance of measuring the presence of inhibitors in PCR templates by using a known positive standard, and provide an approach that will facilitate use of PCR to monitor environmental samples of mosquitoes that harbor endosymbionts or pathogenic organisms.

Introduction:

Wolbachia are obligate intracellular bacteria that cause reproductive distortions such as cytoplasmic incompatibility (CI), parthenogenesis, feminization, and male-killing in the various arthropods they infect (Serbus et al. 2008). *Wolbachia* were first described as pleomorphic, rickettsia-like organisms in Giemsa-stained smears from *Culex pipiens* L. gonads (Hertig 1936). The association of *Wolbachia* with cytoplasmic incompatibility (Yen and Barr 1971) and its potential utility as a genetic drive mechanism to control mosquito populations (Sinkins 2004) have stimulated renewed interest in these bacteria for transgenic mosquito replacement, alteration of population size or age structure, and disruption of pathogen transmission by mosquito vectors. *Wolbachia* infections can be detected by crosses between mosquito strains, fluorescent and electron microscopy (O'Neill et al. 1997), and western blotting (Dobson et al. 1999). The polymerase chain reaction (PCR) (O'Neill et al. 1992, Zhou et al. 1998) has been used for both qualitative and quantitative detection of *Wolbachia,* both in insects and in cell lines (O'Neill et al. 1997).

Although PCR provides a fast and simple method to detect *Wolbachia,* many considerations need to be addressed in the experimental design. For example, false negative reactions with arthropod materials are well-documented (Jeyaprakash and Hoy 2000). With mosquitoes, DNA template has been prepared from different life stages of the insect, and samples range from pools of whole insects to dissected material from individual mosquitoes. Most studies are based on the assumption that recovery of template DNA from biological samples is quantitative, that PCR reactions are uniformly

efficient with each DNA template, and that all reactions remain within the "linear" range of the PCR assay, wherein band intensity is directly correlated with template copy number. These considerations are particularly important in measuring results based on quantitative PCR. For example, in their description of a 20,000-fold range in *Wolbachia* density in a natural population of *Drosophila innubila* Spencer, Unckless et al. (2009) effectively controlled for variability of PCR amplification efficiency by using serial dilutions.

Here we show that an inhibitor that produces false-negative PCR reactions is found in the head of *Culex pipiens* mosquitoes. False-negatives can be eliminated by decapitating the mosquitoes before DNA extraction. This precaution substantially reduced PCR variability among individuals in an infected colony, and facilitated reliable discrimination between infected and antibiotic-cured individuals. In further studies, we found evidence for a PCR inhibitor in four of six additional insect species surveyed, including the mosquito *Anopheles punctipennis* Say. Detection of potential PCR inhibitors by using simple PCR-based assays incorporating known standards will provide a useful tool for monitoring the efficacy of *Wolbachia*-based strategies for control of vector populations, as well as for monitoring pathogen transmission and ecology of endosymbionts in native and transgenic mosquito populations.

Materials and Methods:

Mosquitoes. *Culex pipiens* larvae from the Buckeye strain, collected in Columbus, OH and established in colony in 2000 (Robich and Denlinger 2005) were

obtained from D. Denlinger, Department of Entomology, Ohio State University, in June 2006. Mosquitoes were maintained at 25°C with a photoperiod of 16:8 (L:D) h. Bloodmeals were provided on hamsters (University of Minnesota IACUC Protocol No. 1002A77232), anesthetized with 20% isoflurane in 1, 2-propanediol (Itah et al. 2004). From these wild type mosquitoes, we derived a cured, *Wolbachia*-free "sister colony" over a period of 4 mo, essentially as described by Potaro and Barr (1975). Briefly, we transferred 2-d egg rafts to distilled water containing tetracycline (12.5 μg/ml) and larval food (*Escherichia coli* and Kordon [Hayward, CA] fish fry food), and reared larvae from 10 to 20 egg masses in 3 liters of distilled water in the continuous presence of antibiotic. Recovery of larvae from egg masses decreased during the course of tetracycline treatment. Adults were blood-fed, and their offspring were maintained for two generations (designated G1 and G2) in the absence of tetracycline. Larvae from the G2 adults were maintained for four successive generations in the presence of tetracycline, and subsequent generations of cured mosquitoes were reared in the absence of tetracycline. Loss of *Wolbachia* was monitored by PCR using DNA extracted from individual mosquitoes, and by microscopic observations. With the infected Buckeye strain, we never observed egg rafts that were negative for *Wolbachia* by PCR (N, ≈20 egg rafts), nor did we observe ovaries or testes that failed to contain bacteria-like particles (N, ≈100 individual dissections).

Other Insects. Crickets [*Acheta domesticus* (L.)], and mealworms (*Tenebrio molitor* L.) were from un-characterized laboratory colonies; face flies (*Musca autumnalis* De Geer) were obtained from R. Moon, Department of Entomology, University of

Minnesota; and honey bees (*Apis mellifera* L.), from M. Spivak, Department of Entomology, University of Minnesota. *Drosophila melanogaster* Meigen were from M. O'Connor, Department of Genetics, Cell Biology and Development, University of Minnesota. *Anopheles punctipennis* were reared from larvae collected in Afton, MN. DNA extractions were as described for *Cx. pipiens*.

DNA Extraction. DNA was extracted as described by Livak (1984). Whole or decapitated mosquitoes were individually homogenized in 200 μl of 120-mM Tris-HCl, pH 9, containing 0.5% SDS, 80-mM NaCl, 160-mM sucrose, and 60-mM EDTA. After 30 min at 65°C, potassium acetate (28 μl of 8 M) was added, mixed by vortexing, and the sample was incubated on ice for 30 min. Samples were centrifuged for 10 min in a microcentrifuge at 13,000 rpm, and the resulting supernatant (180 μl) was placed into a new 1.5-ml microcentrifuge tube and 360 μl of 100% ethanol was added. The samples were then briefly vortexed and held overnight at −20°C. Nucleic acids were pelleted by centrifugation at 13,000 rpm for 10 min and the pellets were dried under vacuum, before resuspension in 10-mM Tris-HCl, pH 7.5, containing 0.4-M NaCl and 10-μg boiled RNAseA (400 μl) at 37°C for 1 h. Samples were extracted with an equal volume of phenol, and the aqueous phase (380 μl) was transferred to a new microcentrifuge tube. The phenol phase was re-extracted with 400 μl of 10-mM Tris-HCl, pH 7.5, containing 0.4-M NaCl, and the combined aqueous phases were precipitated with ethanol overnight at −20°C. DNA was recovered by centrifugation, washed in 70% ethanol, dried, and dissolved in 100-μl double-distilled water by sonication in a cup-horn Misonix ultrasonic

liquid processor (Qsonica LLC., Newton, CT) at 90 mA for 30-s intervals, over a total time of 7 min.

Polymerase Chain Reaction. *Wolbachia* primers were based on the genes of ribosomal proteins rpS12 (*rps*L) and rpS7 (*rps*G), which are encoded by adjacent genes in "str operon" as previously described (Fallon 2008). The PCR reaction (20 μl) contained 2.5-mM magnesium chloride, each of the four deoxy-ribonucleotide triphosphates at 0.20 mM; primers at 400 nM; Promega Go-*Taq* polymerase (2.5 U per reaction; Promega, Madison, WI); and 1–9 μl of template DNA. The forward primer was 5′-GCACTAAGGTGTATACTACAACTCC, and the reverse primer was 5′-GCCTTATTAGCTTCAGCCAT. PCR was carried out for 35 cycles with a denaturing step at 95°C for 1 min, annealing at 56°C for 1 min, followed by extension at 72°C for 1 min with a final extension at 72°C for 3 min. The strain designation based on the *Tr1* gene (Duron et al. 2005) was based on sequence obtained with PCR primers F4N: 5′-GCCAAGTGCGTGTATAGTTGAC and R1N: 5′-ATGGAGCTGAAGGTATAGAGG as described above, using an annealing temperature of 59°C. PCR products were electrophoresed on 2% agarose gels and photographed with UV light illumination. Images were "inverted" electronically to show dark bands on a white background. DNA sequencing was carried out at the University of Minnesota BioMedical Genomics Center.

Results:

Derivation of Wolbachia-Free Cx. pipiens. *Culex pipiens* (Buckeye strain) were provided 10% sucrose and blood-fed on hamsters. Egg rafts were collected for isolation

of a *Wolbachia*-free sister colony, using tetracycline treatment at 12.5 μg/ml by using the modified technique of Potaro and Barr (1975). Egg hatch was poor during the first five generations of selection, presumably reflecting cytoplasmic incompatibility within random sib-matings because of loss of *Wolbachia* at varying rates among individual larvae, negative effect of tetracycline treatment on larvae or their microbial diet, or both. After six generations, cured and infected lines exhibited similar larval growth rates, egg hatch, and developmental time. Crosses between males from the infected line and females from the cured line failed to produce offspring.

Variability of Wolbachia PCR Detection. *Wolbachia* infection status of individual mosquitoes yielded unpredictable results when template DNA from whole mosquitoes was amplified by PCR using *Wolbachia*-specific primers. In particular, by PCR, our wild type (infected) colony appeared to contain a mixed population of infected and uninfected males and females (Fig. 1, lanes 3–18). To measure whether inconsistent recovery of DNA caused this variability, we labeled *w*AlbB-infected Aa23 cells (Fallon 2008) with ^{3}H[thymidine], and monitored radioactivity throughout our DNA extraction. Consistent recovery of labeled DNA suggested that the variability in the PCR reactions was not caused by random loss of template. Likewise, to ensure that DNA was uniformly distributed in our samples, we sonicated each sample of purified DNA as a final step in our extraction. Occasionally, sonication would revert a false-negative to positive, suggesting that on occasion, the DNA pellet was not completely dissolved, but in most cases the sonication did not affect PCR results.

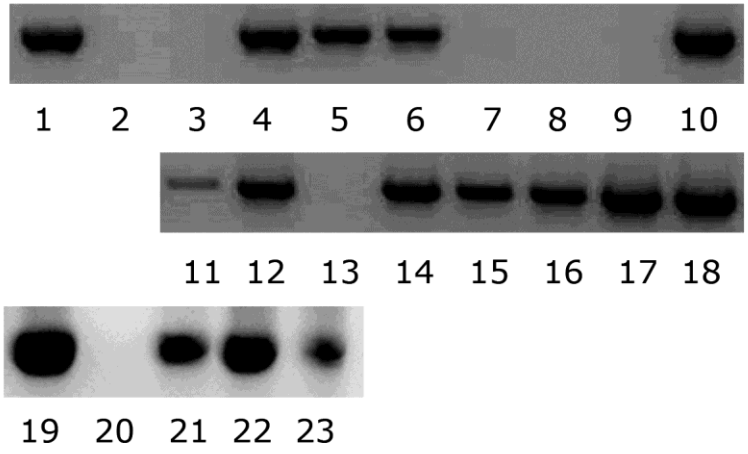

1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18
19 20 21 22 23

Fig 1. Variable PCR-based detection of *Wolbachia* in adult *Cx. pipiens*. Lanes 1 and 19: positive control; lanes 2 and 20: negative control; lanes 3–10 are females, and lanes 11–18 are males. DNA was extracted from individual mosquitoes as described in the Materials and Methods. Lanes 21–23 show positive PCR identification of *Wolbachia* in egg rafts: lane 21, one egg mass; lane 22, pool of five egg masses; lane 23, pool of 10 egg masses.

In some assays, PCR detection in both males and females consistently yielded 100% false-negative results and it appeared that our colony had completely lost the *Wolbachia* infection. In contrast, PCR results with DNA from egg rafts were always positive (Fig. 1, lanes 21–23). Microscopic examination with the cell-permeant dye, Syto-13 indicated that ovaries from our wild type colony released a halo of bacteria-like particles under hypotonic staining conditions, and contained intracellular bacteria-like particles, while ovaries from the cured strain lacked these particles. These experiments, as well as our observation of typical CI in egg rafts resulting from matings between wild type males and antibiotic cured females, showed that our wild type colony was uniformly infected by *Wolbachia,* which was in conflict with the PCR results from whole mosquito templates.

Evidence for a PCR Inhibitor. To measure whether purified mosquito DNA contained an inhibitor, we used a dilution series of DNA template in the PCR reaction. When the sample volume was reduced to 1 μl, we recovered a strong PCR band (Fig. 2, lane 3), and this template continued to give a positive PCR product with up to 10,000-fold further dilution (not shown). In contrast, using 2 μl of the original template substantially reduced the intensity of the positive band (Fig. 2, lane 4), and larger volumes of template DNA failed to produce a PCR product (Fig. 2, lanes 5–11). To further establish the presence of an inhibitor, we tested whether the purified whole mosquito DNA inhibited the PCR reaction of a known positive control (Fig. 2, lanes 12–18). When mixed with increasing concentrations of whole mosquito DNA extract, the

PCR band from positive control DNA progressively declined (Fig. 2, compare lanes 14–18).

In additional studies, we eliminated the possibility that the PCR inhibitor was an artifact of the Livak (1984) procedure, and noted that the inhibitor persisted when we prepared template with a Qiagen DNA kit (Qiagen, Valencia, CA; data not shown) developed for stool samples. We also checked whether the inhibitor was originating specifically from *Wolbachia,* but mosquito DNA purified from infected and uninfected individuals caused comparable levels of inhibition. Interestingly, we noted that DNA pellets commonly had a pink tinge, and reasoned that this pigment might derive from the eyes. Decapitating mosquitoes before homogenization eliminated the inhibitor (Fig. 3, lanes 4–9) whereas DNA extracted from the entire mosquito required a ninefold dilution to yield a positive PCR band (Fig. 3, compare lanes 10, 12, and 14 (1-μl template) with lanes 11, 13, and 15 (9-μl template). Before discovery of the PCR inhibitor, only 53 out of 197 *Culex pipiens* mosquitoes from our infected colony were shown to be infected with *Wolbachia* by PCR. After including decapitation in our DNA extraction protocol, 69 out of 69 *Culex pipiens* from the same infected colony tested positive for *Wolbachia.*

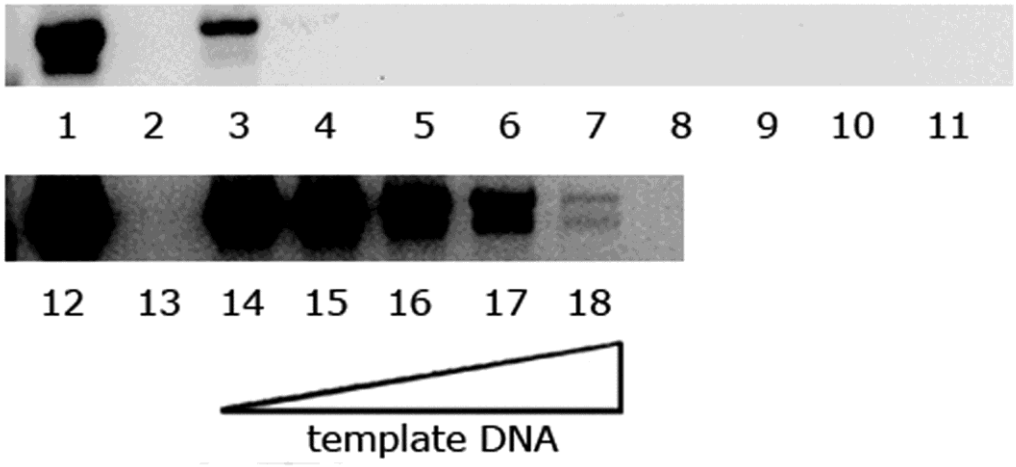

1 2 3 4 5 6 7 8 9 10 11
12 13 14 15 16 17 18
template DNA

Fig 2. Effect of template volume on the PCR reaction. Lanes 1 and 12 show positive controls; lanes 2 and 13 are negative controls; lanes 3–11: 1 μl to 9 μl of DNA template, respectively. Lanes 14–18 all contain positive control DNA as in lane 12, with no additional mosquito DNA (lane 14) and 1 μl to 4 μl of mosquito template DNA (lanes 15–18, respectively).

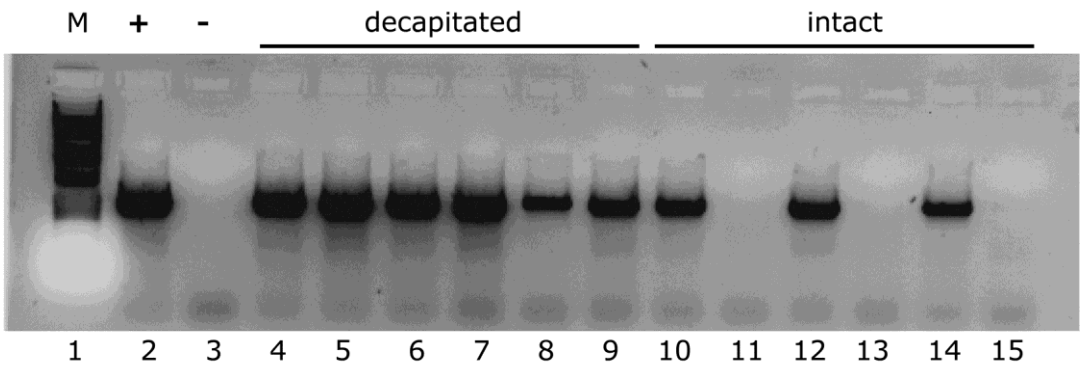

M
+
-
decapitated
intact
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Fig 3. Preparation of template DNA from decapitated mosquitoes removes the inhibitor. Lanes 1, 2, and 3 show DNA ladder, positive control and negative control, respectively. Lanes 4–9 show PCR template DNA prepared from decapitated mosquitoes. For lanes 10–15, mosquitoes were homogenized intact. Even lanes used 1 μl of template DNA; odd lanes had 9 μl of template DNA.

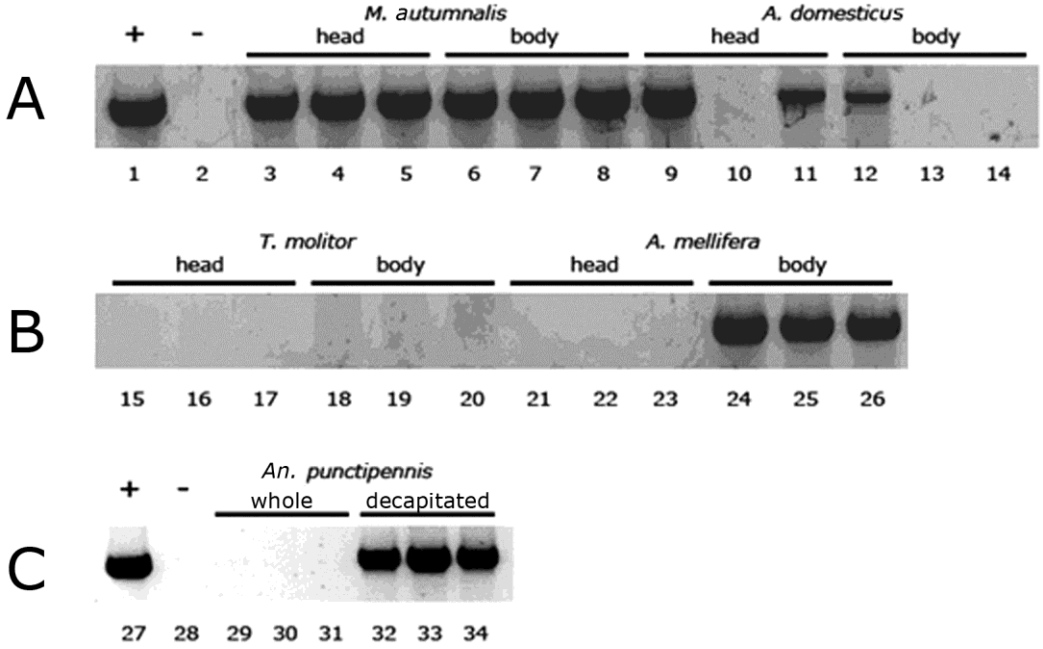

A
M. autumnalis
A. domesticus
+
-
head
body
head
body
1 2 3 4 5 6 7 8 9 10 11 12 13 14
B
T. molitor
A. mellifera
head
body
head
body
15 16 17 18 19 20 21 22 23 24 25 26
C
An. punctipennis
+
-
whole
decapitated
27 28 29 30 31 32 33 34

Fig 4. Survey for PCR inhibitors from various insects. Positive control DNA (1 μl; lanes 1 and 27) was combined with 8-μl template DNA from various extractions (lanes 3–34). Lanes with reduced PCR product, relative to the positive control show evidence for an inhibitor. Panels A and B show lanes from the same gel, whereas Panel C is from a separate gel. Lanes 2 and 28 are negative controls without DNA. Lanes 3–5 show *M. autumnalis* head DNA and 6–8 show *M. autumnalis* decapitated, whole body DNA. Lanes 9–11 show *A. domesticus* head DNA and 12–14 show *A. domesticus* decapitated, whole body DNA. Lanes 15–17 show *T. molitor* head DNA and 18–20 contain *T. molitor* decapitated, whole body DNA. Lanes 21–23 show *A. mellifera* head DNA and 24–26 show *A. mellifera* decapitated, whole body DNA. Lanes 29–30 are *An. punctipennis* whole mosquitoes and 32–34 are decapitated *An. punctipennis* mosquitoes.

Other Insects. To test the prevalence of PCR inhibitors in other insects, we conducted PCR assays measuring inhibition in the separated head and body of miscellaneous insects, including the mosquito, *An. punctipennis*(Fig. 4). We detected an inhibitor within the heads of *A. mellifera* (Fig. 4, lanes 21–23) and *An. punctipennis* (Fig. 4, lanes 29–31) but not in the body of these insects (Fig. 4, lanes 24–26; 32–34). PCR inhibitor(s) were detected in both the head and body of *A. domesticus,* in which levels of inhibitor were particularly variable between individuals, and seemed to be somewhat higher in the body, relative to the head (Fig. 4, lanes 9–14). With DNA from *T. molitor,* inhibitor was present in both heads and body (Fig. 4, lanes 15–20), but we found no evidence for inhibitors in either head or body samples from *M. autumnalis* (Fig. 4, lanes 3–8) and *D. melanogaster* (data not shown).

W. pipientis Strain Characterization. Based on DNA sequence analysis of the transposable element *Tr1,* Duron et al. (2005) identified five *Wolbachia* strains, and showed that North American populations of *Culex pipiens* are singly infected with *w*Pip1 (Florida) or *w*Pip4 (California), or doubly infected with both *w*Pip1 and *w*Pip4 (Minnesota). We sequenced PCR products amplified with primers F4N and R1N from three individual mosquitoes in both directions, and found complete identity with the *w*Pip1 sequence (GenBank accession no. AJ646884) reported by Duron et al. (2005). Thus, based on the *Tr1* gene, the Buckeye strain that originated from Ohio has the same *Wolbachia* genotype as the Florida population described by Duron et al. (2005).

Discussion:

Although Laven (1967) pioneered use of *Wolbachia*-mediated CI to reduce vector populations >40 yr ago, symbiont-based strategies for mosquito population replacement are only recently enjoying renewed attention, due in part to advances in molecular technologies that allow relatively simple detection of *Wolbachia* and exploration of its effects in insect hosts. Of particular interest are recent reports that *Wolbachia* can be successfully transferred into mosquitoes that are uninfected in nature (Xi et al. 2005), that *Wolbachia* can be used to suppress dengue transmission (Hoffmann et al. 2011), and that *Wolbachia* inhibits development of the malaria parasite *Plasmodium* through stimulation of the mosquito immune system (Moreira et al. 2009, Hughes et al. 2011). Despite these remarkable advances, few investigators are investigating the *Wolbachia* infection in natural mosquito hosts, such as *Cx. pipiens*.

Cx. pipiens populations worldwide are infected with *Wolbachia,* and at least five *Wolbachia* strains can be distinguished by sequence analysis of the *Tr1* gene, which encodes a transposable element (Duron et al. 2005). Only two *Wolbachia* strains have been described in North American populations, and strain *w*Pip1 in the recently-colonized Buckeye population of *Cx. pipiens* from Ohio is among these two. As we continued to monitor the *Wolbachia* infection in the wild type mosquitoes, relative to that in a cured sister colony derived by antibiotic treatment, we were puzzled by PCR results that suggested an unstable *Wolbachia* infection in the Buckeye mosquito population.

Spontaneous loss of *Wolbachia* in *Culex* colonies has not to our knowledge been reported, and despite negative PCR results, our colony continued to exhibit microscopic evidence for infection. These considerations supported our suspicion that the PCR results were in error.

While surveying the presence of *Wolbachia* in diverse insects, Jeyaprakash et al. (2000) noted false negative PCR results, and suggested a modification called "long PCR," in which two different polymerases were used simultaneously. Noda et al. (2001) suspected an inhibitor while comparing *Wolbachia* titers in two planthopper species. In one, *Laodelphax striatellus* Fallen, *Wobachia* detection seemed to be consistent and accurate, but in the other, *Sogatella furcifera* (Horváth) these researchers had problems detecting *Wolbachia* in adult males and unsuccessfully tested for an inhibitor by running a dilution series. By measuring PCR band intensity with mosquito DNA template prepared from whole and decapitated mosquitoes and amplified in the presence of an internal positive control, we showed that mosquito heads contain an inhibitor of the PCR reaction. Preparation of template DNA by using a commercially available kit failed to remove the inhibitor, whose molecular identity remains unknown. Inhibition of PCR reactions with DNA extracts from vector mosquitoes is a cause for concern because extracts from the head and thorax are often expected to be enriched for pathogens (Vezzani et al. 2011), whose presence could be masked by the inhibitor.

We included the honey bee in our survey for PCR inhibitors, because most honey bee pathogens are diagnosed by PCR. For example, Chen et al. (2006) investigated transmission dynamics of deformed wing virus (DWV) by PCR assays on dissected

tissues. Virus-positive samples were detected in every tissue, including feces, hemolymph, gut, ovaries, spermatheca, and eviscerated body, but not in the honey bee head. Similarly, Yue and Genersch (2005) detected DWV in the thorax and abdomen of symptomatic and asymptomatic bees, but never detected viral RNA in heads except in symptomatic bees where viral titers were extreme. Although these results suggested that DWV cannot replicate in head tissues, Zioni et al. (2011) recently showed that a recombinant form of DWV does replicate in the honey bee head, suggesting that in at least some studies, others have unknowingly encountered an inhibitor of the PCR reaction in honey bee heads. We note that in studies with *Plasmodium,* the presence of inhibitors from mosquitoes interfered with detection of low parasite numbers (Schriefer et al. 1991, Arez et al. 2000).

In the absence of appropriate positive controls for the PCR reaction, qualitative differences in the abundance of *Wolbachia* under different conditions can be difficult to measure, as PCR inhibition could, for example, mimic a low bacterial load. Echaubard et al. (2010) used quantitative PCR to investigate whether the *Wolbachia* load in a population of insecticide resistant *Cx. pipiens* mosquitoes changed, relative to measurements 36 generations earlier (Berticat et al. 2002). An apparent decrease in *Wolbachia* density in insecticide-resistant mosquitoes, both in the lab and in the field, was attributed to attenuation of the *Wolbachia* infection in the insecticide resistant strains. Given the apparent variability of *Wolbachia* density with diverse factors such as host and *Wolbachia* genotype, environment, age, larval density, and other variables that may be difficult to control (Unckless et al. 2009), results based on quantitative PCR could

be strengthened by incorporating additional controls with internal standards, and showing that the quantitative results 'add up' as expected. Such an approach might lead to a better understanding of *Wolbachia's* effects on host physiology and fitness.

Acknowledgements:

This work was supported by NIH grant AI081322 and by the University of Minnesota Agricultural Experiment Station, St. Paul, MN. We thank Cassandra Kurtz for help with mosquito rearing, and G. D. Baldridge for helpful discussions.

CHAPTER 3:

DETECTION OF THE *WOLBACHIA*-ENCODED DNA BINDING PROTEIN, HU beta, IN MOSQUITO GONADS

Preface to Chapter 3. The following chapter is a reproduction of a manuscript published in the journal Insect Biochemistry and Molecular Biology, Mar 2013; 43(3): 272–279 authored by John F. Beckmann, Todd W. Markowski, Bruce A. Witthuhn, and Ann M. Fallon. Chapter 2. described initial studies developing and characterizing two mosquito colonies which were infected and tetracycline-cured of *Wolbachia* infection. That research was conducted so that I could begin proteomic comparisons of infected vs. uninfected insect tissues. The research contained within this chapter details the initial studies and results from experiments comparing proteomic extracts of infected/uninfected testes and ovaries from the *Culex* colonies. Although the theories I expounded in the discussion of this chapter were later replaced by new ones supported by new data, the research is still relevant because it was the first study to show that *Wolbachia* proteins could be detected and verified by tandem mass spectrometry within infected reproductive tissues. In this light, this publication was a proof of principle, which led to further more insightful studies.

Summary:

Wolbachia are obligate intracellular bacteria that cause cytoplasmic incompatibility in mosquitoes. In an incompatible cross, eggs of uninfected females fail to hatch when fertilized by sperm from infected males. We used polyacrylamide gel electrophoresis and tandem mass spectrometry to identify *Wolbachia* proteins in infected mosquito gonads. These included surface proteins with masses of 25 and 18 kDa and the DNA binding protein, HU, which potentially plays a role in cytoplasmic incompatibility. Using reverse transcriptase polymerase chain reaction, we showed that the HU gene is transcribed in *Wolbachia*-infected *Culex pipiens* and *Aedes albopictus* mosquitoes. We sequenced HU genes from four *Wolbachia* strains and compared deduced protein sequences with additional homologs from the databases. Among the *Rickettsiales*, *Wolbachia* HU has distinct N- and C-terminal basic/acidic amino acid motifs as well as a pair of conserved, cysteine residues. Similarities to eukaryotic architectural chromatin proteins underscore a potential role for HU in cytoplasmic incompatibility.

HIGHLIGHTS

- *Wolbachia* express a DNA binding protein, HU, in mosquito testes and ovaries.
- HU abundance is comparable to that of *Wolbachia* surface proteins.
- *Wolbachia* HU has unique amino acid motifs absent in *E. coli* homologs.
- HU is the first detected *Wolbachia*-encoded protein that potentially binds sperm DNA.

Introduction:

Hertig (1936) first described *Wolbachia* as pleomorphic rods and coccoid bodies in stained gonad smears of *Culex pipiens* mosquitoes. These Gram-negative, alpha proteobacteria, now known as *Wolbachia pipientis*, are classified in the family *Anaplasmataceae*, order *Rickettsiales*, and infect many orders of insects. *Wolbachia* manipulate and distort insect reproduction, causing cytoplasmic incompatibility (CI), parthenogenesis, feminization, and male-killing (Serbus et al., 2008). These reproductive distortions skew offspring ratios in a way that provides a reproductive advantage to females infected by *Wolbachia*, facilitating the organism's spread through naïve insect populations. *Wolbachia*-mediated CI represents a unique tool for manipulating mosquito populations, with potential applications such as vector life shortening (McMeniman et al., 2009), gene drive and population replacement (Rasgon et al., 2006) and reduction of vector competence (Frentiu et al., 2010). An understanding of the molecular basis for CI would represent an important advance towards these goals.

In mosquitoes, CI is a form of conditional sterility that occurs when sperm from *Wolbachia*-infected males fertilize eggs from uninfected females. CI is thought to result from a *Wolbachia*-mediated modification on developing spermatocytes/spermatids (Clark et al., 2003) that can be rescued when the male pronucleus matures in the cytoplasm of eggs from infected females. In eggs from uninfected females, sperm modification in the absence of the rescue factor disrupts cell cycle synchrony between male and female pro-nuclei (Serbus et al., 2008; Callaini et al., 1997); in diploid insects such as mosquitoes, eggs from a CI cross fail to hatch.

Although molecules that mediate CI have not been discovered, several lines of evidence are consistent with the hypothesis that *Wolbachia* secrete one or more effector protein(s) that associates with sperm DNA and interferes with male pronuclear chromatin architecture. Presgraves (2000) showed that the CI effect in *Drosophila* originates from a modification on paternal chromatin. Landmann et al. (2009) then showed that CI in *Drosophila* is associated with impaired ability to deposit maternal histones on male pronuclear chromatin. High CI expression correlates with high *Wolbachia* load in testes (Clark et al., 2003; Clark et al., 2002), which would conceivably raise the concentration of an effector protein, making CI more potent. Increased copulation lowers CI rates, implying depletion of an effector molecule as new sperm develop (Karr et al., 1998). *Wolbachia* genomes encode all of the components of the bacterial type IV secretion system (T4SS), which mediates extracellular export of proteins and DNA (Rances et al., 2005). Moreover, in sperm and ovarian tissues *Wolbachia* often localize around the nucleus and directly contact the nuclear envelope, consistent with the possibility that they secrete molecules into the nucleus (Ferree et al., 2005; Clark et al., 2002; 2003).

The link between CI potency and *Wolbachia* density argues against the secretion or activation of a signaling molecule or transcription factor needed at low concentrations. More likely would be secretion of a protein whose global concentration directly causes CI; such a protein might be expressed at high enough levels to be detected by mass spectrometry, so long as one could acquire enough tissue with a high *Wolbachia* infection. An argument against a secreted effector protein is the observation that in *Nasonia vitripennis*, *Wolbachia* do not need to be present in the germline to induce CI

(Clark et al., 2008). However, this point is countered by the fact that in such cases, *Wolbachia* heavily infect the somatic cyst cells surrounding the developing sperm. In *Drosophila*, these cyst cells are linked directly to the developing sperm by gap junctions that potentially mediate passage of chemical effector molecules (Kiger et al., 2000; Tazuke et al., 2002).

Despite the possibility of a proteomic basis for CI, mass spectrometry-based approaches to study of *Wolbachia* infections have been underutilized. Using SDS PAGE and tandem mass spectrometry we identified protein bands with masses of approximately 25 and 18 kDa that were present in gonads of *Culex pipiens* mosquitoes infected with *Wolbachia pipientis, wPip*, but absent in uninfected tissues. Among smaller protein bands, we detected a *Wolbachia* protein (gi|190571020) homologous to the DNA-binding protein, HU beta, in *Escherichia coli.* *Wolbachia* HU beta is transcribed in both *Culex pipiens* and *Aedes albopictus* infected mosquitoes, further showing that the protein is present in-vivo and that it is expressed by both *w*Pip and *w*AlbA/B strains of *Wolbachia*.

Materials and Methods:

Mosquitoes and other insects. Colonies of *Culex pipiens pipiens* (Buckeye strain) mosquitoes were maintained at 25°C as described previously (Beckmann and Fallon, 2012). Larvae were fed pulverized rat chow and yeast. Adults were allowed to feed on 10% sucrose in water. *Cx. pipiens* mosquitoes are naturally infected with *w*Pip. A cured colony of mosquitoes was established by tetracycline treatment. Infection status was verified by PCR as detailed previously (Beckmann and Fallon, 2012). *Aedes albopictus*

mosquitoes (Houston strain, doubly infected with *w*AlbA and *w*AlbB) were generously provided by Dr. S. L. Dobson (University of Kentucky). Bedbugs (*Cimex lectularius*) were provided by Dr. S. Kells (Department of Entomology, University of Minnesota) and their *Wolbachia* infection is designated *w*Lec. *Wolbachia* from the planthopper *Laodelphax striatellus* (*w*Str) originated from an infected *Ae. albopictus* AeAl2 cell line (Noda, 2002).

Protein extraction. Testes (150) or ovaries (30) were dissected in 100% ethanol and collected in a 1.5 ml tube filled with 100% ethanol, which prevented tissues from sticking to the metal dissecting tools. Pooled tissues were sonicated at 40 mA for 10 seconds in a Kontes GE 70.1 ultrasonic processor, and trichloroacetic acid (TCA) was added to a final concentration of 10% (v/v). After centrifugation at 13,000 rpm in a microcentrifuge, the resulting pellets were washed with acetone:water (9:1), dried, and stored at -20°C.

SDS PAGE and mass spectrometry. Protein samples were reconstituted in SDS sample buffer and boiled prior to electrophoresis, which was usually conducted on 8-18% gradient polyacrylamide gels. Protein gels were submitted to the University of Minnesota's Center for Mass Spectrometry and Proteomics for gel staining with Deep Purple (GE Healthcare), imaging, and in-gel trypsin digestion as described by Anderson et al. (2010). Tryptic peptides were rehydrated in water/acetonitrile (ACN)/formic acid (FA) 98:2:0.1 and loaded using a Paradigm AS1 autosampler system (Michrom Bioresources, Inc., Auburn, CA). Each sample was subjected to Paradigm Platinum Peptide Nanotrap (Michrom Bioresources, Inc.) pre-column (0.15×50 mm, 400-µl

volume) followed by an analytical capillary column (100 μm×12 cm) packed with C18 resin (5 μm, 200 Å MagicC18AG, Michrom Bioresources, Inc.) at a flow rate of 250 nl/min. Peptides were fractionated on a 60 min (10– 40% ACN) gradient on a MS4 flow splitter (Michrom Bioresources, Inc.).

Mass spectrometry (MS) was performed on an LTQ (Thermo Electron Corp., San Jose, CA). Ionized peptides eluting from the capillary column were subjected to an ionizing voltage (2.0 kV) and selected for MS/MS using a data-dependent procedure alternating between an MS scan followed by five MS/MS scans for the five most abundant precursor ions. Tandem mass spectra were extracted by Sequest (Thermo Fisher Scientific, San Jose, CA, USA; version 27, rev. 12. Charge state deconvolution and deisotoping were not performed. All MS/MS samples were analyzed using Sequest (Thermo Fisher Scientific, San Jose, CA, USA; version 27, rev. 12). Sequest was set up to search an rs_wolbachia_aedes_v200808_cRAP_flavivirusREV database (containing protein entries from sequenced *Wolbachia* genomes, the *Aedes aegypti* genome, and flavivirus genomes available as of July, 2011, 74570 entries) assuming the digestion enzyme trypsin and specifying two missed trypsin cleavage sites and one non-tryptic terminus. Sequest was searched with a fragment ion mass tolerance of 0.80 Da and a parent ion tolerance of 1.00 Da. Iodoacetamide derivative of cysteine was specified in Sequest as a fixed modification, and oxidation of methionine was specified as a variable modification. Scaffold (version Scaffold_3.6.0, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0%

probability as specified by the Peptide Prophet algorithm (Keller et al., 2002). Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

DNA extraction and Polymerase Chain Reaction (PCR). DNA was extracted from decapitated mosquitoes (Beckmann and Fallon, 2012) as described by Livak (1984). Infection status was monitored by PCR amplification with primers S12F and S7R as detailed previously (Beckmann and Fallon, 2012); *hup*B sequences were produced with primers HuCloneF: 5' TGGGAATTCGAACAATATTAAGGTAATTTATGAG (near *w*Pip *rps*I gene at 664557-664578 of the *Culex quinquefasciatus* Pel *w*Pip genome) and HuCloneR: 5' TGGGAATTCGAACGAGGCTATATTTCATGGC (in *w*Pip *pdx*J at 664923-664941 of the *Culex quinquefasciatus* Pel *w*Pip genome; underlined bases correspond to a *Bst*BI restriction enzyme site added for cloning purposes). After an initial denaturation at 94 °C for 5 min, DNA was denatured at 94°C for 1 min, annealed at 52°C for 1 min, and extended at 72°C for 1 min for 35 cycles with a final extension at 72°C for 3 min. PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. PCR products were sequenced at the University of Minnesota Biomedical Genomics Center.

RNA extraction, DNase treatment and Reverse Transcriptase (RT) PCR. Cultured cells were pelleted by centrifugation at 800 rpm for 10 minutes, washed in

phosphate-buffered saline, pelleted again by centrifugation at 800 rpm for 10 minutes and resuspended in ice cold lysis buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$, 0.5% (v/v) Nonidet P-40). Pools of 15 decapitated mosquitoes were frozen in liquid nitrogen and ground to powder. The powder was resuspended in 375 µl ice cold lysis buffer and held on ice for 5 minutes. Particulate material was pelleted by centrifugation at 4°C, 13,000 rpm and the supernatant was placed into a new tube. SDS (4 µl of a 20% stock) was added (a final concentration of 0.2%) and immediately mixed into the extract. Proteinase K (2.5 µl of 20 mg/ml stock) was added (a final concentration of 120 µg/ml) and incubated at 37°C for 15 minutes. RNA was then extracted twice with 400 µl phenol/chloroform/isoamyl alcohol (25:24:1) and once with chloroform/isoamyl (24:1). The aqueous phase was recovered, and 40 µl of 3 M sodium acetate, pH 5.2 was added, followed by 2 volumes of 100% ethanol. A precipitate was allowed to form overnight at -20°C. Pellets were washed in 70% ethanol, dried, and resuspended in water. Immediately prior to RT-PCR, samples were treated with Promega's RQ1 RNase-Free DNase (catalog # M610A), according to the manufacturer's instructions (Promega Corporation Madison, WI.) RT-PCR was carried out as described in the Applied Biosystems GeneAmp RNA PCR kit (catalog # N808-0017) with slight modifications. The initial annealing step for the reverse transcriptase reactions was done at 50°C for 5 min, the extension step was at 42°C for 1 h, and the reaction was terminated by heating at 99°C for 5 min. The primer used in the reverse transcriptase reaction to make cDNA was HuMosR and HuR (see below). PCR reactions had an initial denaturation at 94°C for 5 min, then 35 cycles of 94°C denaturing for 1 min, 50°C, 56°C, or 65°C annealing for 1

min, and extension at 72°C for 1 min. Primer pairs used in the PCR reactions were the *wPip* specific HU primers, HuF: 5' AGGATCAGCTAAGTCGCAAAGGCG and HuR: 5' ACCCTTGTCTTTTCAGGAACGGTC (56°C annealing), the *Wolbachia* HU primers designed to be conserved within *Wolbachia* that infect mosquitoes, HuMosF: 5' ATGAGTAAAGAAGATATARTAAAC and HuMosR: 5' TCATAATCTCACCATTTTGAG, (50°C annealing) and the mosquito ribosomal protein S3 (RpS3) primers 40S3F: 5' ATGCCGAGAAGGTCGCCAC and 40S3R: 5' GCACGGATCTCCGGAATGG (65°C annealing).

Sequence Alignments. Experimentally obtained DNA sequences were translated using the ExPASy translate tool, http://web.expasy.org/translate/, from the SIB Swiss Institute of Bioinformatics. Amino acid sequence alignments were constructed by using MUSCLE in the MEGA 5.05 software program. Alignments were performed under default settings with Gap penalties: Gap open -2.9, Gap Extend 0, Hydrophobicity Multiplier 1.2, and Memory/Iterations: Max Memory in MB 1686, Max Iteration 8 (Edgar, 2004). Sequence data from the *Aedes albopictus* sample potentially containing both *w*AlbB and *w*AlbA DNA was identical to the NCBI translated protein Reference Sequence: ZP_09542731.1 from *w*AlbB. Sequence data from *w*Pip was identical to the published sequence, GenBank: AM999887.1; sequences from *w*Str and *w*Lec are deposited under GenBank accession numbers 000000 and 000000, respectively.

Results:

Wolbachia membrane proteins in infected gonads. SDS polyacrylamide gradient gels were used to compare proteins in testes and ovaries from infected and tetracycline-cured *Cx. pipiens* mosquitoes (Fig 1). The similar pattern of stained proteins in infected and cured tissues indicated that in mosquito reproductive tissues, *Wolbachia* infection is accompanied by little overall change in proteins detectable by visual inspection of stained polyacrylamide gels. Despite the overall similarity, bands at 25 and 18 kDa, unique to infected testes and ovaries, were observed in a series of 5 replicate experiments with independent pools of dissected tissues (Fig 1A). Additional bands with masses below 14 kDa were also typical of infected testes (Fig 1B) and ovaries (not shown). We began analysis by excising the 25 kDa and 18 kDa bands from gels, subjecting tryptic peptides to mass spectrometry, and assessing the presence of *Wolbachia* encoded products.

The most abundant protein in the 25 kDa band was the *Wolbachia* surface protein, WSP. In five separate isolates, we detected a total of 48 peptides covering 61% of the total protein (Table 1 and Fig. 2A). Not surprisingly, a smaller number of peptides from host proteins were also recovered from the 25 kDa band, including the ~ 25 kDa proteasome subunits and ubiquitin, previously found to be up-regulated when cultured cells were newly infected with *w*AlbB (Fallon and Witthuhn, 2009). The 18 kDa band contained 41 peptides covering 45% of a "putative *Wolbachia* membrane protein" (Figs. 1A and 2B). Detection of this membrane protein was repeated in two biological replicates. In aggregate, these data reveal that abundantly expressed *Wolbachia* surface

proteins can be identified by gel electrophoresis and mass spectrometry against a background of host proteins.

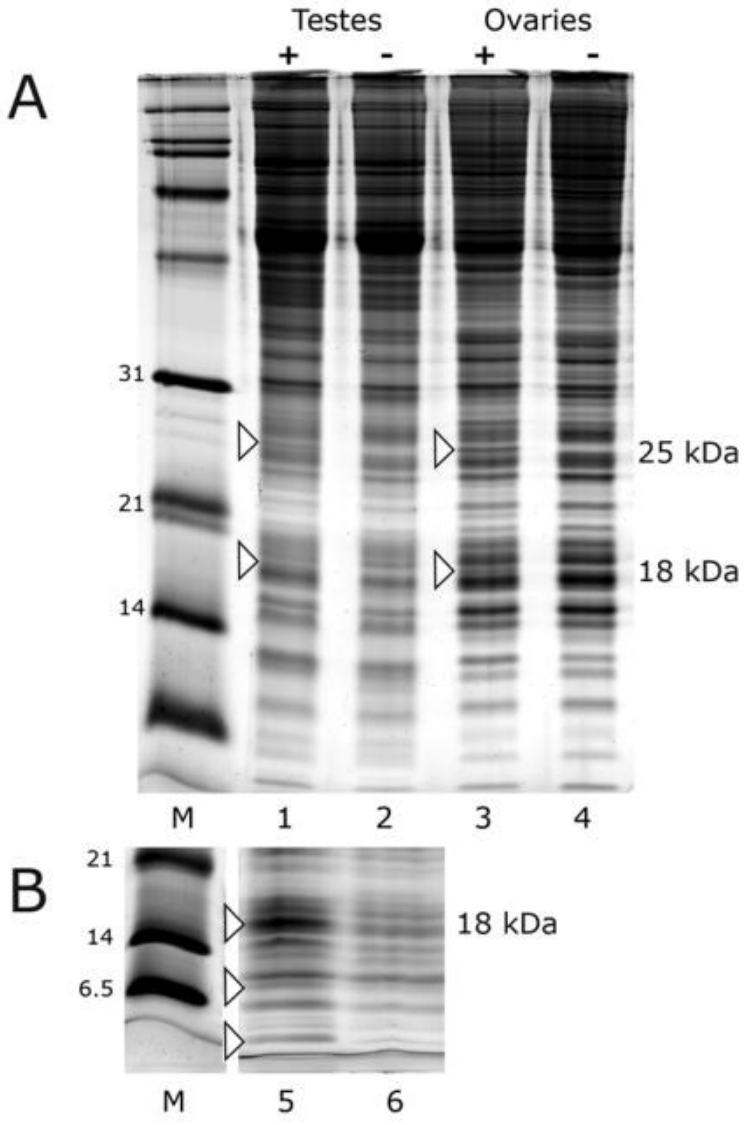


Fig 1. Electrophoretic analysis of proteins extracted from *Cx. pipiens* gonads. Panel A. Lanes 1 and 2 are 150 individual testes from *Cx. pipiens*; lanes 3 and 4 are 30 individual ovaries. Lanes 1 and 3 (+) are from *Wolbachia*-infected mosquitoes; lanes 2 and 4 are from a tetracycline-cured sister colony. Panel B represents an independent biological replicate from testes of infected (lane 5) and cured (lane 6) *Cx. pipiens*. Protein bands at 25 kDa and 18 kDa were visible in 8 biological replicates. Lane 5 shows a better replicate of the 18 kDa band as well as smaller bands that were examined by mass spectrometry. proteins can be identified by gel electrophoresis and mass spectrometry against a background of host proteins.

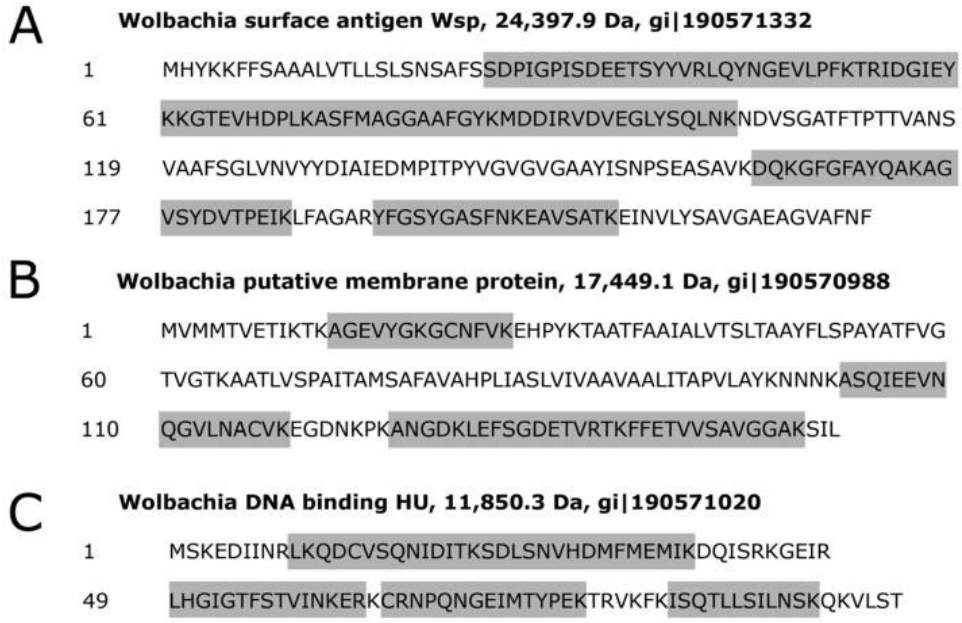

A
Wolbachia surface antigen Wsp, 24,397.9 Da, gi|190571332
1 MHYKKFFSAAALVTLLSLSNSAFSSDPIGPISDEETSYYVRLQYNGEVLPFKTRIDGIEY
61 KKGTEVHDPLKASFMAGGAAFGYKMDDIRVDVEGLYSQLNKNDVSGATFTPTTVANS
119 VAAFSGLVNVYYDIAIEDMPITPYVGVGVGAAYISNPSEASAVKDQKGFGFAYQAKAG
177 VSYDVTPEIKLFAGARYFGSYGASFNKEAVSATKEINVLYSAVGAEAGVAFNF
B
Wolbachia putative membrane protein, 17,449.1 Da, gi|190570988
1 MVMMTVETIKTKAGEVYGKGCNFVKEHPYKTAATFAAIALVTSLTAAYFLSPAYATFVG
60 TVGTKAATLVSPAITAMSAFAVAHPLIASLVIVAAVAALITAPVLAYKNNNKASQIEEVN
110 QGVLNACVKEGDNKPKANGDKLEFSGDETVRTKFFETVVSAVGGAKSIL
C
Wolbachia DNA binding HU, 11,850.3 Da, gi|190571020
1 MSKEDIINRLKQDCVSQNIDITKSDLSNVHDMFMEMIKDQISRKGEIR
49 LHGIGTFSTVINKERKCRNPQNGEIMTYPEKTRVKFKISQTLLSILNSKQKVLST

Fig 2. Total peptide coverage detected for the three *Wolbachia* proteins. A. *Wolbachia* surface protein, WSP. Coverage varied from 43% – 61% among 5 replicates. B. *Wolbachia* putative membrane protein. Coverage varied from 17% - 45% among 2 replicates C. *Wolbachia* DNA binding protein HU beta. Coverage varied from 41% - 67% among 5 replicates. Shaded boxes indicate mass spectrometry identified peptides.

Band	extract source	Protein	kDa	(p)	TS	UP	%C	R	Accession
25 kDa	testes, ovaries, CAS	*Wolbachia* surface protein	25	100%	48	16	61%	5	gi|190571332
18 kDa	testes, ovaries	*Wolbachia* putative membrane protein	17	100%	41	10	45%	2	gi|190570988
<14 kDa	testes, ovaries	*Wolbachia* DNA binding HU	12	100%	9	6	52%	5	gi|190571020

Table 1. *Wolbachia* protein identities obtained from MS/MS analysis of the best individual replicate, Scaffold v3.6. Abbreviations are as follows: (p), identity probability; TS, total spectra matches; UP, unique peptides; %C, % coverage; R, replicates.

A Wolbachia DNA binding protein. Infected tissues differentially expressed bands with masses less than 14 kDa (Fig. 1B) in some samples, but these bands were not consistently replicated in every sample. Likewise, in studies with *Wolbachia*-infected cell lines, we noted that extracts sometimes included differentially expressed, radiolabeled bands with masses below 14 kDa (Fallon et al, 2013). Protein from this region was extracted and subjected to tryptic digestion and mass spectrometry analysis. A high proportion of peptides (9 peptides; 52% coverage, see Table 1 and Fig. 2C) corresponded to the *Wolbachia* DNA binding protein HU (gi|190571020**)**. In subsequent experiments, we identified HU peptides in five biological replicates, including both testes and ovaries, with high confidence and protein coverage ranging from 41% - 67% among replicates.

Sequence and BLAST analysis of HU. BLAST analysis indicated that the *Wolbachia* HU protein was the ortholog of the 90 amino acid protein, HU beta, encoded by *hup*B in *Escherichia coli*, with 23 amino acid identities spanning the length of the protein, and an E value of 1.00e-09 (Fig 3A). The most striking differences between HU beta from *wPip* and *E. coli* were a five amino acid insertion (KQDCV) near the N terminus of the *Wolbachia* protein within alpha helical region α1, and eight residues at the C-terminus extending beyond helical region α3 in the *E. coli* homolog. Alpha helix α1 is the dimerization site that mediates an interaction between HU alpha and HU beta proteins in *E. coli*, but *Wolbachia* genomes lack the *hup*A ortholog, suggesting that in *Wolbachia*, HU beta forms a homodimer, interacts with a unique partner, or functions as a monomer. HU beta from *w*Pip also maintains a conserved intercalating proline residue essential for DNA binding in *E. coli* HU beta (identified by the solid downward pointing

arrow in Fig. 3A), and a second proline (identified in Fig. 3B with an open arrow) was conserved among *Wolbachia* homologs. Just upstream of the conserved proline was the second of a pair of cysteine residues (shaded grey), which potentially forms a disulfide bond with the unique cysteine in the N-terminal KQDCV insertion in *Wolbachia* HU beta.

To determine whether the unique aspects of the N and C termini and the cysteine residues were conserved among *Wolbachia* HU beta homologs, we sequenced the *hup*B genes from the *Wolbachia* strains available in our lab and aligned them with annotated sequences from the databases. Internal primers based on *w*Pip *hup*B nucleotide sequences did not reliably produce *hup*B PCR products in other *Wolbachia* strains, suggesting variability among *hup*B homologs. To address this possibility, primers in better conserved regions, including the 135 nt intergenic region downstream of the gene encoding ribosomal protein S9 (*rps*I; HuCloneF) and sequence within the gene encoding

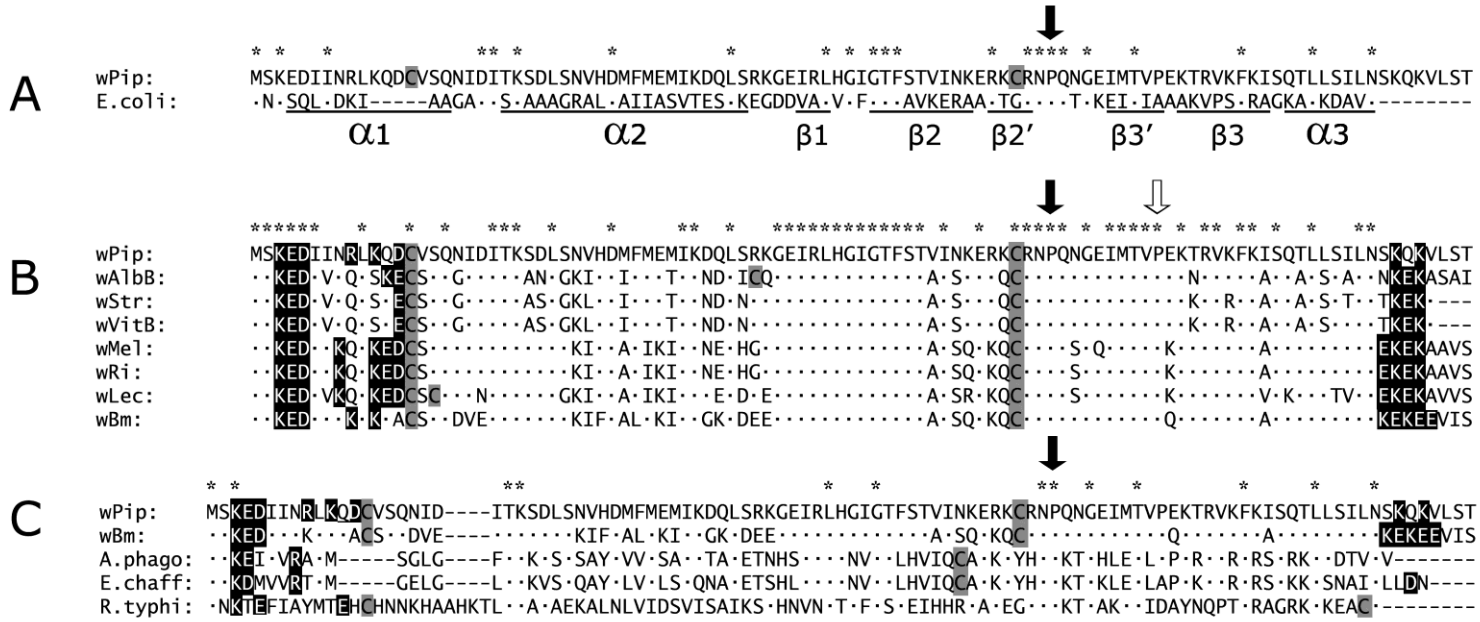

A
wPip: MSKEDIINRLKQDCVSQNIDITKSDLSNVHDMFMEMIKDQLSRKGEIRLHGIGTFSTVINKERKCRNPQNGEIMTVPEKTRVKFKISQTLLSILNSKQKVLST
E.coli: ·N·SQL·DKI-----AAGA··S·AAAGRAL·AIIASVTES·KEGDDVA·V·F···AVKERAA·TG····T·KEI·IAAAKVPS·RAGKA·KDAV·--------
α1 α2 β1 β2 β2′ β3′ β3 α3
B
wPip: MSKEDIINRLKQDCVSQNIDITKSDLSNVHDMFMEMIKDQLSRKGEIRLHGIGTFSTVINKERKCRNPQNGEIMTVPEKTRVKFKISQTLLSILNSKQKVLST
wAlbB: ··KED·V·Q·SKECS··G·····AN·GKI··I···T··ND·ICQ·····A·S···QC·····N·····A··A·S·A··NKEKASAI
wStr: ··KED·V·Q·S·ECS··G·····AS·GKL··I···T··ND·N·····A·S···QC·····K··R··A··A·S·T··TKEK·---
wVitB: ··KED·V·Q·S·ECS··G·····AS·GKL··I···T··ND·N·····A·S···QC·····K··R··A··A·S····TKEK·---
wMel: ··KED··KQ·KEDCS·····KI··A·IKI··NE·HG·····A·SQ·KQC····S·Q·····K·····A·····EKEKAAVS
wRi: ··KED··KQ·KEDCS·····KI··A·IKI··NE·HG·····A·SQ·KQC····S·····K·····A·····EKEKAAVS
wLec: ··KED·VKQ·KEDCSC···N·····GKI··A·IKI···E·D·E·····A·SR·KQC····S·····K·····V·K···TV··EKEKAVVS
wBm: ··KED···K·K·ACS··DVE·····KIF·AL·KI··GK·DEE·····A·SQ·KQC·····Q·····A·····KEKEEVIS
C
wPip: MSKEDIINRLKQDCVSQNID----ITKSDLSNVHDMFMEMIKDQLSRKGEIRLHGIGTFSTVINKERKCRNPQNGEIMTVPEKTRVKFKISQTLLSILNSKQKVLST
wBm: ··KED···K···ACS··DVE----·····KIF·AL·KI··GK·DEE·····A·SQ·KQC·····Q·····A·····KEKEEVIS
A.phago: ··KEI·VRA·M-----SGLG----F··K·S·SAY·VV·SA··TA·ETNHS····NV··LHVIQCA·K·YH··KT·HLE·L·P·R··R·RS·RK··DTV·V-------
E.chaff: ··KDMVVRT·M-----GELG----L··KVS·QAY·LV·LS·QNA·ETSHL····NV··LHVIQCA·K·YH··KT·KLE·LAP·K··R·RS·KK·SNAI·LLDN----
R.typhi: ·NKTEFIAYMTEHCHNNKHAAHKTL··A·AEKALNLVIDSVISAIKS·HNVN·T·F·S·EIHHR·A·EG···KT·AK··IDAYNQPT·RAGRK·KEAC·--------

Fig 3. Comparisons among HU beta proteins. A. Amino acid sequence alignment of *wPip* HU beta and its *E.coli* ortholog. Asterisks indicate amino acid identities, and the downward-pointing arrow indicates a proline residue important in DNA binding. Under the *E. coli* sequence, residues that participate in alpha helix (α) and beta sheet structure (β) as defined by Swinger and Rice (2004) are underlined. B. Alignment of HU beta amino acid sequences. *Wolbachia* genes from strains *w*Pip*, w*AlbB*, w*Lec*, and w*Str were experimentally determined in this study; others were from *Wolbachia* genomes available in the database. C. Amino acid sequence alignment of *Wolbachia* HU beta from (*w*Pip and *w*Bm) compared with representative orthologs from *Anaplasma phagocytophilum*, *Ehrlichia chaffeensis*, and *Rickettsia typhi*. N and C-terminal charged motifs composed of lysine, arginine, glutamic acid, and aspartic acid in HU beta are shown in black boxes. Cysteine residues are shown in grey. For other residues, dots indicate identities.

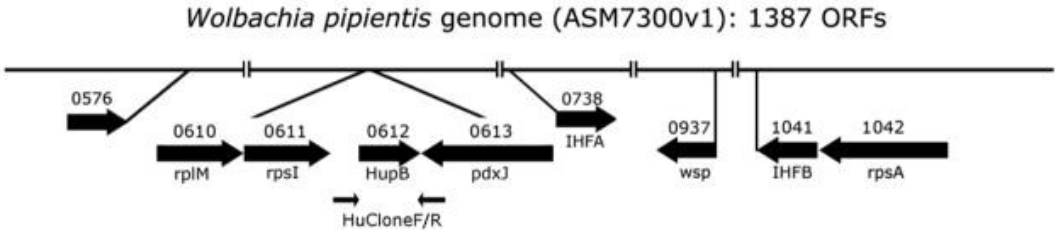

Wolbachia pipientis genome (ASM7300v1): 1387 ORFs
0576
0610
rplM
0611
rpsI
0612
HupB
0613
pdxJ
0738
IHFA
0937
wsp
1041
IHFB
1042
rpsA
HuCloneF/R

Fig 4. Partial linearized genome of *Wolbachia pipientis*. Genes described in this study are mapped by solid arrows showing the direction of transcription. Flanking the *hup*B (0612) are primers HUCloneF/R (indicated by small black arrows), which map immediately downstream of *rps*I and within *pdx*J, respectively, used to clone HU sequences from other *Wolbachia* strains described in Fig. 3. Amino acid sequences of HU, WSP (0937) and the putative membrane protein (0576) are described in Fig. 2. Genes encoding IHF alpha (0738) and IHF beta (1041), transcribed in opposite directions, are also shown. Locations of ribosomal protein genes potentially transcribed as operons that include *hup*B (*rpl*M, *rps*I) and *ihf*B (*rps*A) are also represented.

pyridoxine 5'-phosphate synthase (*pdx*J; HuCloneR, Fig. 4) were used to obtain PCR products encoding *hup*B from *wPip, wStr* (from the planthopper, *Laodelphax striatellus*), *wAlbB* (from the mosquito *Aedes albopictus*), and *wLec* (from the bedbug, *Cimex lectularius*). During our sequence analysis the protein sequences for the *wAlbB* HU beta homolog (ZP_09542731.1) became available. Our translated *wAlbB* protein sequence and our *wPip* HupB DNA sequence (Gene ID: 6385678) matched published sequences perfectly. Sequences for *wStr hup*B and *wLec hup*B were deposited in the GenBank database under accession numbers JX984572 and JX984573.

Alignment of *Wolbachia* HU beta proteins showed 52 amino acid identities common to all available homologs (see the asterisks at the top of the alignment in Fig. 3B), including the N-terminal KED motif, the two internal cysteines, two conserved prolines, and KEK motifs at the C-terminus. Strains *w*AlbB and *w*Lec have an additional third cysteine, but their positions are not conserved. The C-terminal charged motifs appear to be characteristic of *Wolbachia* HU beta proteins, relative to homologs in the other genera in the *Rickettsiales*: *Anaplasma*, *Ehrlichia*, and *Rickettsia* (Fig 3C), but at least one cysteine residue occurred in each of the HU beta representatives of the order *Rickettsiales* aligned in Fig. 3C.

***Expression of Wolbachia hupB*.** We verified that *hup*B is transcribed in two species of infected mosquitoes by reverse-transcriptase PCR (RT-PCR). The *hup*B transcript was detected in RNA prepared from *Wolbachia*-infected male and female *Cx. pipiens* and *Ae. albopictus* mosquitoes (Fig. 5, lanes 3, 5, 7, 9), but not in uninfected cell cultures (Fig. 5, lane 1), indicating that this gene is transcribed from the *w*Pip and

*w*AlbA/B genomes, *in vivo,* at levels detectable in total RNA prepared from whole mosquitoes. RT-PCR bands were excised and sequenced to confirm that the PCR product encoded the *hup*B transcript. Positive controls for RNA quality were performed by amplifying a PCR product from mosquito ribosomal protein RpS3 gene transcripts (Fig. 5, lanes 11-15). Negative controls were uninfected mosquito cells (lane 1) and RNA assayed without reverse transcriptase (lanes 2, 4, 6, 8, 10).

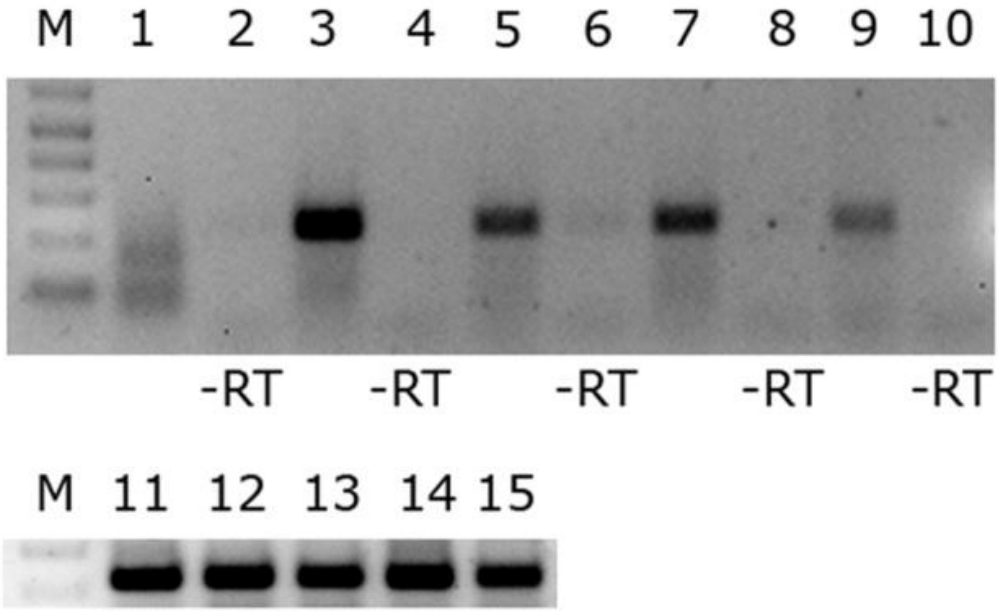

M 1 2 3 4 5 6 7 8 9 10
-RT -RT -RT -RT -RT
M 11 12 13 14 15

Figure 5. RT PCR analysis of the *hup*B transcript in pooled decapitated mosquitoes and uninfected mosquito cells. M is a DNA marker. Lanes 1 and 11 are RNA extracted from uninfected C7-10 mosquito cells. Lanes 3, 4, and 12 are total RNA extracted from *Culex pipiens* females. Lanes 5, 6, and 13 are RNA extracted from *Culex pipiens* males. Lanes 7, 8, and 14 are RNA extracted from *Aedes albopictus* females. Lanes 9, 10, and 15 are RNA extracted from *Aedes albopictus* males. Lanes 2, 4, 6, 8, and 10 were assayed without reverse transcriptase as a control for DNA contamination of RNA preparations. Lanes 11-15 are positive controls for RNA quality using primers for the mosquito ribosomal protein RpS3.

In silico analysis of HU partners. In *E. coli*, HU proteins are a constitutive component of the bacterial nucleoid with roles in replication, transposition, gene inversion, and expression (Nash and Robertson, 1981; Ryan et al., 2002; and Azam et al., 1999). HU binds bacterial DNA and wraps it into a nucleosome-like particle providing compaction and protection. In *E. coli*, HU beta is a member of a group of four nucleoid proteins: two HU proteins, HU alpha and HU beta and two IHF (integration host factor) proteins, IHF alpha and IHF beta, each of which can form homo- and hetero- dimers (α/α, β/β, α/β), but HU proteins do not dimerize with IHF proteins. As described above, all *Wolbachia* genomes lack a gene corresponding to *E. coli hup*A. However, *Wolbachia* genomes encode three HU proteins: gi|190571020, which aligns most closely with *E. coli* HU beta (Fig 3); gi|190571140, which most closely resembles *E coli* IHF alpha, and gi|190571428, which is the homolog of *E. coli* IHF beta. Relative to *E. coli*, *w*Pip IHF alpha had 34% identity over 99 residues, uniformly distributed over the length of the protein (Fig. 6A), and among strains of *Wolbachia*, IHFA was highly conserved (Fig. 6B). Relative to *E. coli*, *w*Pip IHF beta had 28% identity over 92 residues (Fig. 6C), with strong identity among *Wolbachia* strains (Fig. 6D). We note that overall, among *Wolbachia* homologs, IHF alpha appears to be evolving more slowly than IHF beta and both IHF proteins are evolving more slowly than HU beta.

Finally, to evaluate possibilities for co-expression and/or potential expression of these genes as part of an operon(s), we examined the relative locations of the *Wolbachia* genes encoding the proteins that we have detected in this study (Fig. 4). The membrane protein encoding genes, hypothetical protein (wpa 0576) and wsp (wpa 0937) are encoded by distant genes, and transcribed in opposite directions. Interestingly, both *hup*B and *ihf*B might be included in separate operons, each of which includes upstream ribosomal protein genes; strong constitutive expression would be expected if they were co-transcribed with components of the translational machinery. *hup*B is flanked on the 5' end by *rpl*M, encoding 50S ribosomal protein L13, and *rps*I, encoding 30S ribosomal protein S9; *ihf*B is immediately downstream of *rps*A, encoding the 30S ribosomal protein S1. Conservation of the relative order of these genes in *E. coli* is consistent with possible inclusion of *hup*B and *ihf*B proteins in operons that encode key components of the translational machinery.

Discussion:

Detection of Wolbachia proteins. Although CI was successfully used for mosquito population replacement more than 40 years ago (Laven, 1967), the molecular basis for CI remains elusive. We have initiated proteomic studies addressing the hypothesis that in infected mosquito testes, *Wolbachia* secrete effector molecules that interact with sperm DNA. Here we use mass spectrometry to identify three proteins abundantly expressed in *Wolbachia*-infected tissues: two membrane proteins and a DNA binding protein. In addition to WSP, previously identified in *Drosophila* reproductive

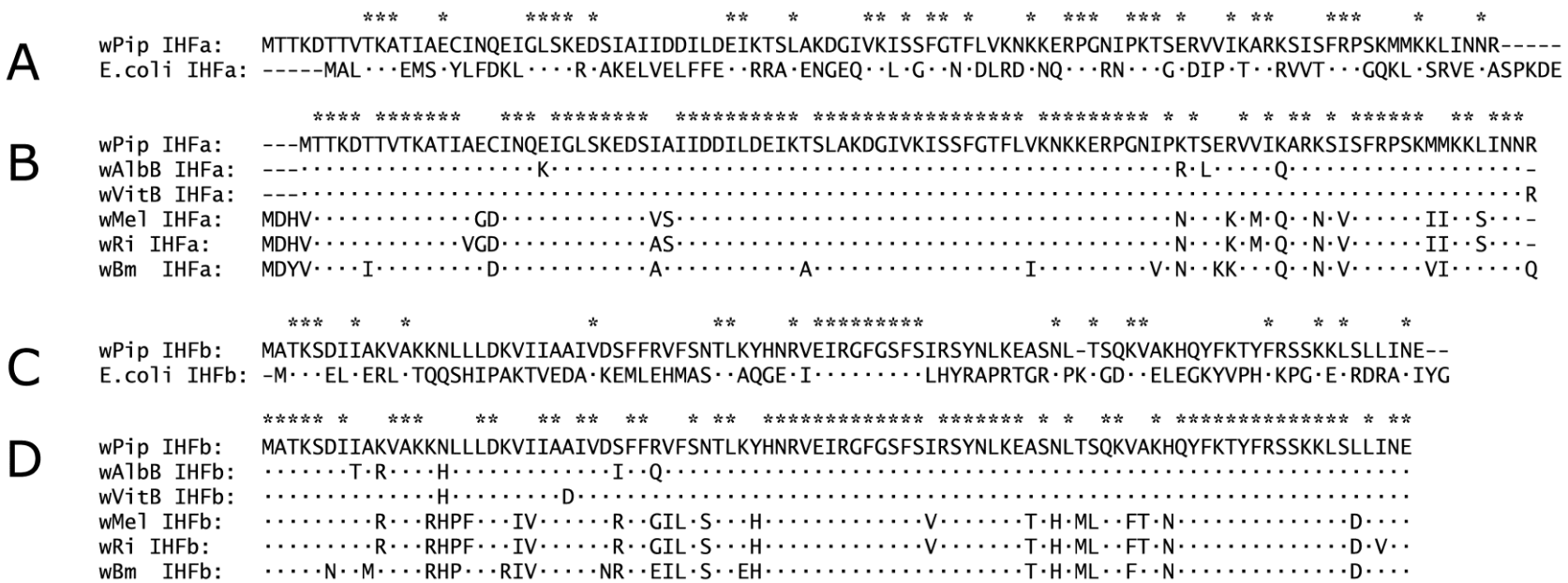

A
wPip IHFa:
E.coli IHFa:
B
wPip IHFa:
wAlbB IHFa:
wVitB IHFa:
wMel IHFa:
wRi IHFa:
wBm IHFa:
C
wPip IHFb:
E.coli IHFb:
D
wPip IHFb:
wAlbB IHFb:
wVitB IHFb:
wMel IHFb:
wRi IHFb:
wBm IHFb:

Fig 6. IHF amino acid sequence alignments. A. Sequence alignment of *wPip* IHF alpha with its *E.coli* ortholog. B. Sequence alignment of known *Wolbachia* IHF alpha homologs. C. Sequence alignment of *wPip* IHF beta with its *E.coli* ortholog. D. Sequence alignment of known *Wolbachia* IHF beta homologs. Asterisks indicate amino acid identities. Identities relative to the *w*Pip proteins are indicated by dots.

tissues (Sasaki et al., 1998; Braig et al., 1998), we also detected an 18 kDa band on SDS gels that contained peptides from a second *Wolbachia* membrane protein of unknown function. Although detection of abundant membrane proteins was not surprising, it remained to be seen whether mass spectrometry was capable of detecting less abundant *Wolbachia* proteins. To optimize collection of sufficient tissue, we dissected mosquitoes in 100% ethanol, rather than buffered saline, to reduce tissue adherence to dissecting instruments. Improved recovery of protein allowed us to detect bands that migrated at masses of less than 14 kDa. *Wolbachia* peptides recovered from this region of the gel corresponded to those of a DNA binding protein homologous to HU beta in *E. coli*.

Structure and function of HU. The three dimensional structure of *E.coli* HU beta is composed of three alpha helices, with five beta sheets between the second and third helix. Alpha 1 helix acts as the dimerization site, and DNA binding involves the five beta sheets and a conserved intercalating proline. In dimer form, two beta sheet arms interact with the minor groove of DNA and force it to bend 180 degrees around the base of the protein (Swinger and Rice, 2004). Note that *Wolbachia* and *E. coli* HU beta proteins share the strongest concentration of identities in this key DNA binding region, including the intercalating proline present in all *Wolbachia* homologs (Fig. 3).

The most striking differences in primary structure of the *Wolbachia* homologs, relative to HU beta in *E. coli*, are the acidic/basic motifs near the N and C termini, precisely where charged protein signals often mediate nuclear localization (Assier et al., 1999), and are thought to be an important signal for *Wolbachia's* type IV secretion system (Vergunst et al., 2005). Moreover, because these motifs interrupt and flank

amino acids that constitute *E. coli* helices α1 and α3, they might affect higher order structure and protein function. These changes are particularly intriguing because all known *Wolbachia* genomes have lost the partner gene *hup*A, and therefore can form only homodimers. Interestingly, other researchers have reported unusual histone1-like PAAK and KAAK additions on the C-terminal domain of a mycobacterial HU protein which lowers its binding constant and makes its interaction with DNA more specific (Kumar et al., 2010). Moreover, the pair of cysteines in *Wolbachia* HU beta potentially form intra or intermolecular disulfide bonds that would potentially decrease the distance between acidic/basic "KEKE-like" motifs, which are thought to be involved in assembly of protein complexes by a putative "charged zipper" mechanism (Realini et al., 1994).

In *Chlamydia* and *E. coli*, HU proteins have been shown to regulate transitions in growth cycle (Wagar and Stephens, 1988; Azam et al., 1999) and to regulate transposition and recombination (Friedman, 1988), which could have profound impacts on genetic manipulation of *Wolbachia*. The relative importance of HU in *Wolbachia* biology is underscored by the fact that of the seven major bacterial nucleoid proteins listed for *E. coli* (Dorman 2009), the *Wolbachia pipientis* genome encodes only three; HU beta and the two IHF proteins described above. Because *Wolbachia* lacks most of the nucleoid proteins known from *E. coli* (Azam et al., 1999; Dorman, 2009) HU beta is likely to play a key role in the *Wolbachia* cell cycle. HU has limited homology to chromosome partitioning protein, MUKB, reviewed in (Nasmyth and Haering, 2005), a 170 kDa prokaryotic member of the SMC (structural maintenance of chromatin) protein family involved in condensation and segregation of chromosomes. In *E. coli*, HU and

MUKB cooperate in chromosome segregation. Double *hup*A/*hup*B mutants are defective in chromosome partitioning and show the same phenotype as *muk*B mutants (Jaffe et al., 1997); that is they develop a population of anucleate cells. Interestingly, *hup*A/*muk*B double mutants are lethal, but having one gene seems to compensate for the loss of the other. The absence of *muk*B and *hup*A homologs in *Wolbachia* genomes further underscores the uniqueness of *Woblachia* HU beta, which presumably accomplishes chromosome segregation in the absence of MUKB.

Of the three nucleoid proteins encoded by *Wolbachia* genomes, we note that both HU beta and IHF beta genes lie immediately downstream of ribosomal protein genes, raising the possibility that they could be co-transcribed with highly expressed components of the translational machinery. However, if HU beta and IHF beta are co-expressed, it is curious that that we did not detect peptides corresponding to *Wolbachia* IHF proteins, which would be expected to migrate with masses similar to that of HU beta. Recovery of several peptides from HU beta, relative to the absence of peptides from IHF beta, leads us to speculate that HU beta may be expressed at higher levels than the IHF proteins, or alternatively, may be more stable, compatible with a possible role for HU beta in mediating CI. We also note that during preparation of this manuscript Darby et al. (2012) published a global study of transcription in *Wolbachia* from the nematode, *Onchocerca ochengi*, and noted up-regulation of an HU transcript in germ tissue as compared to somatic tissues.

HU as a candidate CI effector molecule. If the acidic/basic motifs at the N terminus of *Wolbachia* HU beta are recognized by the T4SS, we envision that HU beta

could be accumulated in the host cell nucleus and bind to sperm DNA, at the time when histones are being replaced by protamines. In *Drosophila* it is claimed that any small positively charged/basic protein can assume the protamine packaging function (Hennig, 2003). All *Wolbachia* HU beta proteins have cysteines that do not occur in *E coli* HU proteins; cysteine residues are known to be essential for protamine-mediated DNA compaction in eukaryotes (Ballhorn, 2007). The possibility that HU beta interacts with DNA in mosquito sperm is further supported by structural and functional similarity between bacterial HU proteins and eukaryotic chromatin architectural HMG (High Mobility Group) box proteins (Oberto et al., 1994; Bianchi, 1994) known to be important for nuclear condensation and chromatin structure in early *Drosophila* embryos (Ner and Travers, 1994). In particular, HMG box proteins specifically favor cruciform DNA. Likewise, In *E. coli*, HU beta homodimers (the only form possible in *Wolbachia*) bind preferentially to cruciform DNA, while heterodimers associate with both linear and cruciform DNA (Pinson et al., 1999). An enrichment of *Wolbachia*-derived homodimeric HU beta in the paternal pronucleus conceivably disrupts chromatin changes associated with fertilization by an antagonistic/competitive mechanism with maternal histones and HMG Box proteins, causing the improper chromatin condensation (Landmann et al., 2009; Breeuwer and Werren, 1990) and aberrant synchrony between male and female pronuclei that have been observed in incompatible crosses in *Drosophila simulans* (Callaini et al. 1997). Future research on HU will focus on identifying additional proteins with which it may interact, further characterization of its role in the *Wolbachia* replication cycle, and evaluation of its possible role in cytoplasmic incompatibility.

CHAPTER 4:

DETECTION of the *WOLBACHIA* PROTEIN WPIP0282 in MOSQUITO SPERMATHECAE: IMPLICATIONS for CYTOPLASMIC INCOMPATIBILITY

Preface to Chapter 4. The following chapter is a reproduction of a manuscript published in the journal of Insect Biochemistry and Molecular Biology, 2013 Sep;43(9):867-78 authored by John F. Beckmann and Ann M. Fallon. After establishing infected and uninfected colonies of *Culex pipiens* (Chapter 2) and performing initial proteomic comparison of their infected vs. uninfected reproductive tissues, we detected a *Wolbachia* expressed DNA binding protein abundantly expressed in infected testes (Chapter 3). The *Wolbachia* DNA binding protein HU beta was a prime candidate for further study because it rationally fit the necessary characteristics of a potential CI inducer. In an effort to determine whether the *Wolbachia* derived DNA binding protein HU beta localized to the sperm I performed extensive dissections of spermathecae (an organ that stores insect sperm) and subjected those tissues to tandem mass spectrometry. We did not detect the presence of the *Wolbachia* HU beta. However, we did detect the accumulation of a single *Wolbachia* protein within the sperm filled spermathecae. The details of those studies are described here in Chapter 4.

Summary:

Cytoplasmic incompatibility (CI) is a conditional sterility induced by the bacterium *Wolbachia pipientis* that infects reproductive tissues in many arthropods. Although CI provides a potential tool to control insect vectors of arthropod-borne diseases, the molecular basis for CI induction is unknown. We hypothesized that a *Wolbachia*-encoded, CI-inducing factor would be enriched in sperm recovered from spermathecae of female mosquitoes. Using SDS-PAGE and mass spectrometry, we detected peptides from the 56 kDa hypothetical protein, encoded by *w*Pip_0282, associated with sperm transferred to females by *Wolbachia* infected males. We also detected peptides from the same protein in *Wolbachia* infected ovaries. Homologs of *w*Pip_0282 and the co-transcribed downstream gene, *w*Pip_0283, occur as multiple divergent copies in genomes of CI-inducing strains of *Wolbachia*. The operon is located in a genomic context that includes mobile genetic elements. The absence of *w*Pip_0282 and *w*Pip_0283 homologs from genomes of *Wolbachia* in filarial nematodes, as well as other members of the *Rickettsiales*, suggests a role as a candidate CI effector.

Highlights:

· Peptides encoded by *w*Pip_0282 occur in sperm from dissected mosquito spermathecae

· Homologs of *w*Pip_0282 are only present in CI-inducing *Wolbachia* strains

· *w*Pip_0282 and *w*Pip_0283 comprise a two gene operon that has duplicated and diverged

· *w*Pip_0283 encodes a C-terminal, eukaryotic SUMO protease domain

Introduction:

Wolbachia are Gram-negative, alpha proteobacteria (family *Anaplasmataceae*, order *Rickettsiales*) that infect a high percentage of arthropod species, up to 76% in one survey (Jeyaprakash and Hoy 2000). *Wolbachia* are also mutualistic endosymbionts of the nematodes *Brugia malayi* and *Onchocerca volvulus*, which in humans cause lymphatic filariasis and onchocerciasis, respectively (Taylor et al., 2012). Recently, it has been shown that *Wolbachia* can cause significant immune responses in humans when liberated from microfilariae (Brattig et al., 2004; Shiny et al., 2011; Bazzocchi et al., 2007). In mosquitoes, *Wolbachia* represent a potential means of controlling vector populations because they manipulate and distort reproduction, causing cytoplasmic incompatibility (CI). CI skews offspring ratios in a way that provides a reproductive advantage to females infected by *Wolbachia*, facilitating the organism's spread through naïve insect populations (Serbus et al., 2008) and providing a gene drive system for population replacement (Rasgon et al., 2006). In addition, some *Wolbachia* infections shorten vector life spans (McMeniman et al., 2009), reduce vector competence (Frentiu et al., 2010; Vavre and Charlat 2012), and interfere with immune mechanisms that facilitate maintenance of pathogens (Teixeira et al., 2008). Elucidation of the molecular mechanisms that cause CI would represent an important advance towards use of *Wolbachia* for biological control of insect pest populations and vector-borne disease.

CI occurs when sperm from *Wolbachia*-infected males fertilize eggs from uninfected females. Clark et al (2003) suggest that during development in *Wolbachia*-infected testes, spermatocytes acquire a *Wolbachia* strain-specific modification that can

be rescued if the male pronucleus matures in cytoplasm of eggs infected with the same (compatible) strain of *Wolbachia*. However, if a modified sperm matures in cytoplasm of eggs from uninfected females, the modification cannot be corrected, and developing embryos show disruptions in cell cycle synchrony (Serbus et al., 2008; Callaini et al., 1997); in diploid insects such as mosquitoes, eggs from a CI cross fail to hatch. CI is complex, and in *Culex pipiens* mosquitoes, 17 different cytotypes have been described, with both unidirectional and bidirectional mating incompatibilities among mosquitoes from different geographical locations (Laven, 1967). Thus far, mechanisms by which *Wolbachia* strains cause the incompatibility patterns observed in crossing experiments are unknown, and recent comparisons of *Wolbachia* genomes from geographically isolated strains of *Culex pipiens* from Johannesburg (JHB) and Sri Lanka (Pel) have uncovered few differences that might account for CI (Klasson et al., 2006; Salzberg et al., 2008).

Several lines of evidence are consistent with the hypothesis that *Wolbachia* secrete one or more effector protein(s) that associates with sperm DNA and interferes with male pronuclear chromatin architecture. The streamlined *Wolbachia* genomes have retained genes encoding all of the components for a functional bacterial type IV secretion system (T4SS), which is known to mediate extracellular export of proteins and DNA (Rances et al. 2008). Moreover, in developing sperm and in ovarian tissues, *Wolbachia* have been shown to localize around the nucleus and directly contact the nuclear envelope, consistent with the possibility that they secrete molecules into the nucleus (Ferree et al., 2005; Clark et al., 2002; 2003). Presgraves (2000) demonstrated that in *Drosophila,* CI is induced by an unknown modification on paternal chromatin; more recently, Landmann et

al. (2009) reported impaired maternal histone deposition on male pronuclear chromatin in *Drosophila*. Severity of the CI phenotype is associated with high *Wolbachia* titers in the testes (Clark et al., 2002; 2003; Veneti et al., 2003; 2004), while increased mating decreases CI penetrance within *Drosophila simulans*, implying depletion of an effector chemical that must have time to accumulate in developing sperm (Karr et al., 1998).

We recently undertook extensive proteomic studies with the goal of identifying potential CI inducing protein candidates in mosquito reproductive organs, including the *Wolbachia* DNA binding protein HU beta (Beckmann et al., 2013). To extend these studies, we hypothesized that CI effector proteins might be expressed at levels sufficient for detection by mass spectrometry in mature spermatozoa recovered from the spermathecae of female mosquitoes. Unlike the diverse developmental stages of sperm recovered from dissected testes, mature spermatozoa, transferred to a female during mating and stored in her spermathecae, are free of *Wolbachia* itself, which are discarded in "waste bags" that eliminate excess cytoplasmic material as the spermatids elongate (Serbus et al., 2008). Thus, we reasoned that dissected spermathecae would be enriched for secreted *Wolbachia* proteins accumulated during spermatogenesis and retained by virtue of association with organelles or nuclear DNA, but would not include *Wolbachia* structural proteins, or proteins involved in *Wolbachia* replication and maintenance. Using SDS polyacrylamide gel electrophoresis (SDS PAGE) and mass spectrometry we identified peptides encoded by *w*Pip_0282 in *Culex pipiens* spermathecae as well as in ovarian tissues. In-silico analysis revealed that *w*Pip_0282 is expressed from a two-gene operon, which has undergone duplication and divergence in *Wolbachia* genomes for

which data are available. RT-PCR data confirmed polycistronic translation of the operon and its duplicate homolog as one mRNA. Intriguingly, *w*Pip_0282/0283 operons occur only in insect-associated, CI-inducing *Wolbachia*, and map to genomic regions characterized by mobile genomic elements, ankyrin repeats, and WO phage genes.

Materials and Methods:

Mosquito Colonies and Maintenence. Colonies of *Culex pipiens pipiens* (Buckeye strain) mosquitoes were maintained at 25°C as described previously (Beckmann and Fallon, 2012). Larvae were fed pulverized rat chow and yeast. Adults were allowed to feed on 10% sucrose in water. *C. pipiens* mosquitoes are naturally infected with *w*Pip. A cured colony of mosquitoes was established by tetracycline treatment. Infection status was verified by PCR as detailed previously (Beckmann and Fallon, 2012). *Aedes albopictus* mosquitoes (Houston strain, doubly infected with *w*AlbA and *w*AlbB) were generously provided by Dr. S. L. Dobson (University of Kentucky).

Cell lines. An *Aedes albopictus* cell line infected with *Wolbachia* strain *w*AlbB was a persistently infected subpopulation, TW-2800, derived from the TW-280 cells described previously (Fallon and Witthuhn, 2009).

Protein Extraction. Mosquito protein extracts were prepared as previously described (Beckmann et al. 2013). Spermathecae (2400 lobes, from 800 mosquitoes) or ovaries (30) were dissected in 100% ethanol and collected in a 1.5 ml tube filled with 100% ethanol, which prevented tissues from sticking to the metal dissecting tools. Pooled

tissues were sonicated at 40 mA for 10 seconds in a Kontes GE 70.1 ultrasonic processor, and trichloroacetic acid (TCA) was added to a final concentration of 10% (v/v). After centrifugation at 13,000 rpm in a microcentrifuge, the resulting pellets were washed with acetone:water (9:1), dried, and stored at -20°C. Representative samples of spermathecae were examined by microscopy to confirm that that the mosquitoes had mated and that *Wolbachia* were absent from the spermathecal tissues. A mixture of 50 μM Syto 13 (Invitrogen, Carlsbad, CA) and 50 μM propidium iodide (5 μL) was added to individual spermathecae (which were dissected in PBS). Samples were then crushed with a glass coverslip and viewed with an Olympus IX70 fluorescent microscope.

SDS PAGE and mass spectrometry. Protein samples were reconstituted in SDS sample buffer and boiled prior to electrophoresis, which was usually conducted on 8-18% gradient polyacrylamide gels. Protein gels were submitted to the University of Minnesota's Center for Mass Spectrometry and Proteomics for gel staining with Deep Purple (GE Healthcare), imaging, and in-gel trypsin digestion as described by Anderson et al. (2010) and detailed in Beckmann et al. (2013). Briefly, tryptic peptides were rehydrated in water/acetonitrile, passed through a Paradigm Platinum Peptide Nanotrap (Michrom Bioresources, Inc.) pre-column, followed by an analytical capillary column (100 μm×12 cm) packed with C18 resin (5 μm, 200 Å MagicC18AG, Michrom Bioresources, Inc.) at a flow rate of 250 nl/min. Peptides were fractionated on a 60 min (10– 40% ACN) gradient on a MS4 flow splitter (Michrom Bioresources, Inc.). Mass spectrometry (MS) was performed on an LTQ (Thermo Electron Corp., San Jose, CA) as described previously (Beckmann et al., 2013). Tandem mass spectra were extracted by

Sequest (Thermo Fisher Scientific, San Jose, CA, USA; version 27, rev. 12) and searched against an rs_wolbachia_aedes_v200808_cRAP_flavivirusREV database (containing protein entries from sequenced *Wolbachia* genomes, the *Aedes aegypti* genome, and flavivirus genomes available as of July, 2011, 74570 entries) assuming the digestion enzyme trypsin and specifying two missed trypsin cleavage sites and one non-tryptic terminus. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003) using previously described criteria (Beckmann et al., 2013). As a negative control a decoy-database was searched with all samples resulting in a 0.0% False Discovery Rate under the settings of 3 minimum peptides per protein at a 95% minimum protein threshold.

DNA extraction, Polymerase Chain Reaction (PCR), RNA extraction, DNase treatment and Reverse Transcriptase (RT) PCR. DNA was extracted from decapitated mosquitoes according to Beckmann and Fallon, 2012. RNA extracts were produced from cultured cells and mosquitoes as follows: Cultured cells were pelleted by centrifugation at 800 rpm for 10 minutes, washed in phosphate-buffered saline, pelleted again by centrifugation at 800 rpm for 10 minutes and resuspended in ice cold lysis buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$, 0.5% (v/v) Nonidet P-40). Pools of 15 decapitated mosquitoes were frozen in liquid nitrogen and ground to powder. The powder was resuspended in 375 µl ice cold lysis buffer and held on ice for 5 minutes. Particulate material was pelleted by centrifugation at 4°C, 13,000 rpm and the supernatant was placed into a new tube. SDS (4 µl of a 20% stock) was added (a final concentration of 0.2%) and immediately mixed into the extract. Proteinase K (2.5 µl of 20 mg/ml stock)

was added (a final concentration of 120 µg/ml) and incubated at 37°C for 15 minutes. RNA was then extracted twice with 400 µl phenol/chloroform/isoamyl alcohol (25:24:1) and once with chloroform/isoamyl alcohol (24:1). The aqueous phase was recovered, and 40 µl of 3 M sodium acetate, pH 5.2 was added, followed by 2 volumes of 100% ethanol. A precipitate was allowed to form overnight at -20°C, and recovered by centrifugation. Pellets were washed in 70% ethanol, dried, and resuspended in water. Immediately prior to RT-PCR, samples were treated with Promega's RQ1 RNase-Free DNase (catalog # M610A), according to the manufacturer's instructions (Promega Corporation, Madison, WI.) RT-PCR was carried out as described in the Applied Biosystems GeneAmp RNA PCR kit (catalog # N808-0017) with slight modifications. The initial annealing step for the reverse transcriptase reactions was done at 50°C for 5 min, the extension step was at 42°C for 1 h, and the reaction was terminated by heating at 99°C for 5 min. Reverse primers were used in the initial reverse transcriptase reaction to make cDNA. PCR reactions had an initial denaturation at 94°C for 5 min, then 35 cycles of 94°C denaturing for 1 min, annealing for 1 min (see table S1 for temperatures), and extension at 72°C for 1 min. All primers and primer specifications are listed in Table S1.

Infection status of mosquitoes was monitored by PCR amplification with primers S12F and S7R as detailed previously (Beckmann and Fallon, 2012); PCR and RT-PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. PCR and RT-PCR products were sequenced at the University of Minnesota Biomedical Genomics Center.

Sequence Alignments and BlastP Analysis. Experimentally obtained DNA sequences were translated using the ExPASy translate tool, http://web.expasy.org/translate/, from the SIB Swiss Institute of Bioinformatics. Amino acid sequence alignments were constructed under default settings using clustalW2 from the EMBL-EBI website, http://www.ebi.ac.uk/Tools/msa/clustalw2/. Conserved protein domains were identified using the NCBI's Conserved Domain Database in conjunction with BlastP and also independently verified using EMBL-EBI's InterProScan (Zdobnov and Apweiler, 2001). BlastP analysis was conducted using the NCBI BlastP program.

Results:

Detection of WPIP0282 in Spermathecae and Ovaries. To search for *Wolbachia* proteins in sperm enriched extracts we dissected sperm-filled spermathecae (approximately 2400, from 800 female mosquitoes per lane; each individual mosquito contains 3 round spermathecal lobes that individually store sperm), and total protein was extracted for analysis by SDS-PAGE. Analysis showed two bands that migrated between 45 and 66 kDa in protein from spermathecae filled with *Wolbachia* modified sperm (Fig 1, lane 1), relative to unmodified sperm from tetracycline cured males (Fig 1, lane 2). Similarly, two bands between 45 and 66 kDa were seen in infected ovarian extracts (Fig 1, lane 3), but not in cured ovarian extracts (Fig 1, lane 4). One of these bands appeared to be the same molecular weight in both ovarian and spermathecal extracts, but the others appeared to be of different sizes. Aside from these two unique bands in spermathecal and ovarian tissues, there were few differences in the overall banding pattern of spermathecal,

ovarian, and testicular extracts. This is similar to the case described previously for extracts from mosquito ovaries and testes (Beckmann et al., 2013). Spermathecal extracts also lacked the characteristic 24 kDa band associated with the *Wolbachia* surface protein, consistent with the exclusion of *Wolbachia* from mature sperm; this is in contrast to extracts from *Wolbachia*-infected ovaries (Fig 1, lane 3 white arrow), testes (Beckmann et al., 2013) and infected cultured cells (Fallon et al., 2013).

When analyzed by MS/MS, the bands from *Wolbachia*-modified sperm contained 4 unique peptides from the hypothetical 56 kDa translation product originating from the *w*Pip_0282 gene gi|190570728; (Fig 2; note the four peptides shaded in black; we will refer to the corresponding protein as WPIP0282). Ovarian bands (designated by solid arrows in Fig 1, lane 3) analyzed by MS/MS, also included the same four spermathecal *Wolbachia* peptides and an additional 14 unique peptides derived from the predicted translation product of *w*Pip_0282 in the annotated genome *w*Pip (Pel) endosymbiont of *Culex pipiens quinquefasciatus* (shaded grey in Fig 2). Sequence coverage of WPIP0282 among four independent SDS-PAGE and mass spectrometry experiments ranged from 7-54%, with lower coverage in extracts from spermathecae (7-9%) and higher coverage in ovaries (9-54%), which harbor both intact *Wolbachia* as well as putative *Wolbachia*-secreted proteins. While the magnitude of differential expression based on visual inspection of SDS-PAGE gels varied among biological replicates, WPIP0282 peptides were identified with a protein identity probability of 100% in two independent dissections of spermathecae and two independent dissections of infected ovaries. We note that the protein was detected in multiple bands which themselves migrated at different theoretical

molecular weights in both spermathecae and ovarian extracts. Such banding shifts can be the result of post-translational modifications and/or peptide cleavage. However at this time the physiological cause of these banding patterns is unknown.

Homologs of WPIP0282 in Wolbachia from Culex and Aedes Mosquitoes. We used PCR-based approaches to sequence the wPip_0282 homolog (KF114896) from the Buckeye strain of *Culex pipiens* maintained in our lab (see Beckmann and Fallon, 2012) in order to evaluate identity and compare amino acid sequences between the deduced translation product and WPIP0282 from the genome of *w*Pip (Pel). The deduced protein sequence [designated *w*Pip (Buckeye) in Fig 2] had four amino acid differences over 491 residues (boxed in Fig. 2) relative to the protein encoded by *w*Pip (Pel)_0282 in the annotated genome of *Wolbachia pipientis* Pel (endosymbiont of *Culex pipiens quinquefasciatus* mosquitoes from Sri Lanka; see Klasson et al., 2008; hereafter we refer to both proteins interchangeably as WPIP0282). We then used BlastP analysis to determine the copy number of the gene within the various *w*Pip genomes in *Culex* mosquitoes. The *w*Pip strain JHB (from *Culex pipiens quinquefasciatus* from Johannesburg, South Africa, Salzberg et al., 2008) encodes a full-length homolog of WPIP0282 [491 residues, ZP_03335575.1, designated *w*Pip (JHB)2 in Fig 2] and two partial homologs, one of which [394 residues, ZP_03335681.1, designated *w*Pip (JHB)1 in Fig 2] is nearly complete. The third homologous accession (ZP_03335652.1) in *w*Pip (JHB) has only 67 residues, and will not be discussed further here. Deduced sequences of *w*Pip (JHB) proteins were largely identical to each other, and to WPIP0282 in the Pel and Buckeye strains, with the exception of a block of diversity spanning ~ 20 residues in

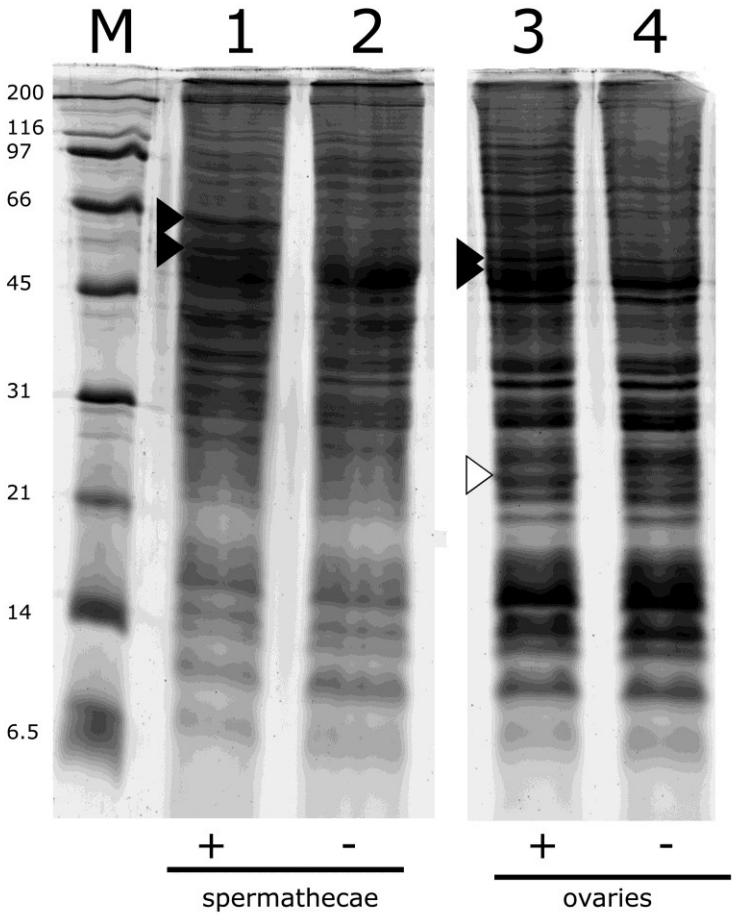

M
1
2
3
4
200
116
97
66
45
31
21
14
6.5
+
-
+
-
spermathecae
ovaries

Figure 1. SDS-PAGE analysis of protein extracted from reproductive tissues from *Wolbachia*- infected and uninfected *Culex pipiens* mosquitoes. Lane M, molecular mass markers (kDa); lanes 1 and 2, 2400 dissected spermathecal lobes; lanes 3 and 4, 30 ovaries. Positive (+) and negative (-) symbols below the lanes indicate tissue derived from *Wolbachia* infected mosquitoes and tetracycline cured mosquitoes, respectively. Black arrows indicate bands containing at least three peptides from the WPIP0282 protein. The white arrow in lane 3 indicates the *Wolbachia* WSP protein band in infected ovaries (see Beckmann et al., 2013) which is absent from spermathecal extracts (lane 1). Microscopic examination of spermathecae was used to confirm that that the mosquitoes had mated and that *Wolbachia* were absent from the spermathecal tissues.

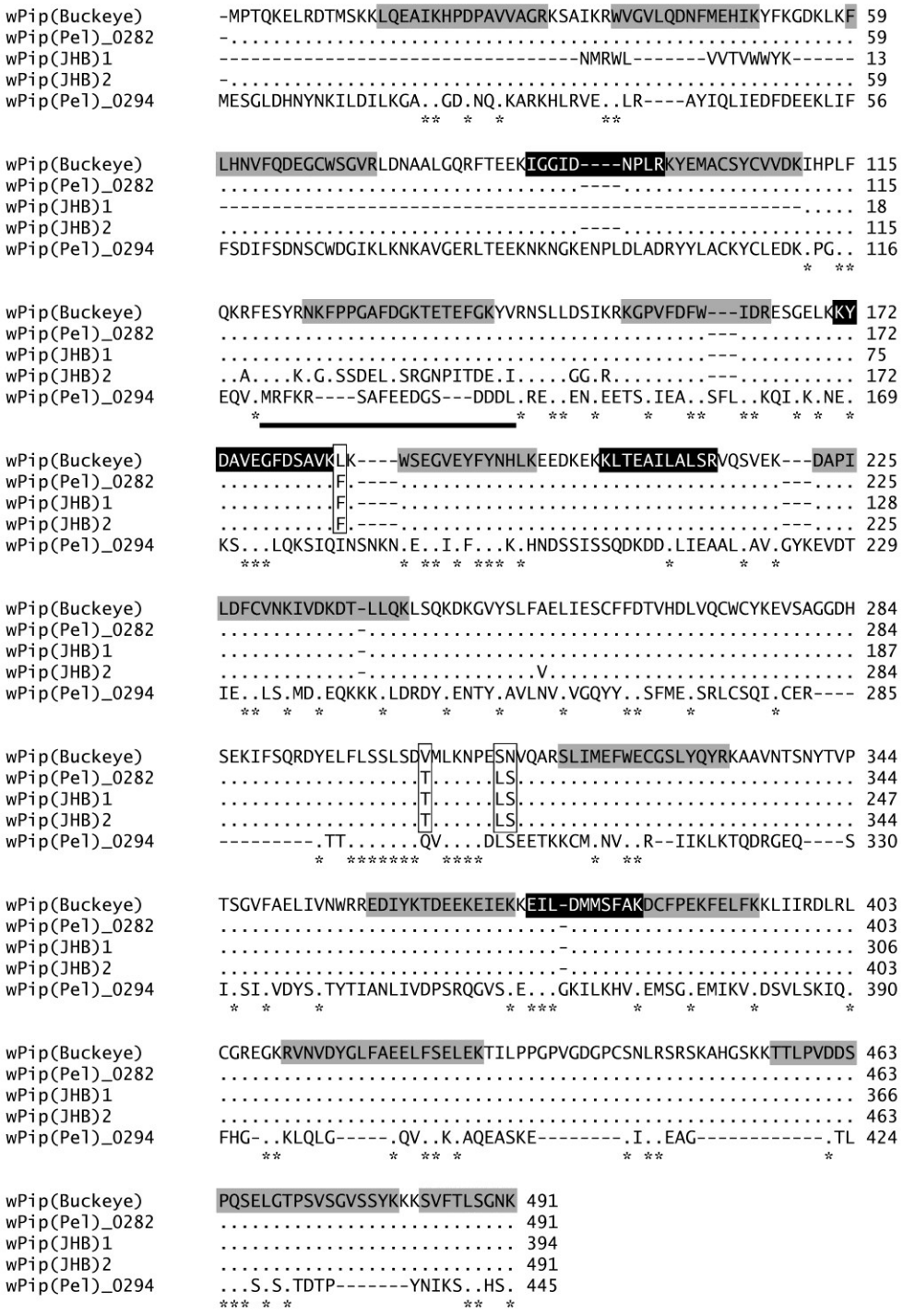

wPip(Buckeye) -MPTQKELRDTMSKKLQEAIKHPDPAVVAGRKSAIKRWVGVLQDNFMEHIKYFKGDKLKF 59
wPip(Pel)_0282 -..... 59
wPip(JHB)1 ----------NMRWL-------VVTVWWYK------ 13
wPip(JHB)2 -..... 59
wPip(Pel)_0294 MESGLDHNYNKILDILKGA..GD.NQ.KARKHLRVE..LR----AYIQLIEDFDEEKLIF 56
wPip(Buckeye) LHNVFQDEGCWSGVRLDNAALGQRFTEEKIGGID----NPLRKYEMACSYCVVDKIHPLF 115
wPip(Pel)_0282 .....----..... 115
wPip(JHB)1 ----------..... 18
wPip(JHB)2 .....----..... 115
wPip(Pel)_0294 FSDIFSDNSCWDGIKLKNKAVGERLTEEKNKNGKENPLDLADRYYLACKYCLEDK.PG.. 116
wPip(Buckeye) QKRFESYRNKFPPGAFDGKTETEFGKYVRNSLLDSIKRKGPVFDFW---IDRESGELKKY 172
wPip(Pel)_0282 .....---..... 172
wPip(JHB)1 .....---..... 75
wPip(JHB)2 ..A....K.G.SSDEL.SRGNPITDE.I.....GG.R.....---..... 172
wPip(Pel)_0294 EQV.MRFKR----SAFEEDGS---DDDL.RE..EN.EETS.IEA..SFL..KQI.K.NE. 169
wPip(Buckeye) DAVEGFDSAVKLK----WSEGVEYFYNHLKEEDKEKKLTEAILALSRVQSVEK---DAPI 225
wPip(Pel)_0282 .....F.----.....---.... 225
wPip(JHB)1 .....F.----.....---.... 128
wPip(JHB)2 .....F.----.....---.... 225
wPip(Pel)_0294 KS...LQKSIQINSNKN.E..I.F...K.HNDSSISSQDKDD.LIEAAL.AV.GYKEVDT 229
wPip(Buckeye) LDFCVNKIVDKDT-LLQKLSQKDKGVYSLFAELIESCFFDTVHDLVQCWCYKEVSAGGDH 284
wPip(Pel)_0282 .....-..... 284
wPip(JHB)1 .....-..... 187
wPip(JHB)2 .....-.....V..... 284
wPip(Pel)_0294 IE..LS.MD.EQKKK.LDRDY.ENTY.AVLNV.VGQYY..SFME.SRLCSQI.CER---- 285
wPip(Buckeye) SEKIFSQRDYELFLSSLSDVMLKNPESNVQARSLIMEFWECGSLYQYRKAAVNTSNYTVP 344
wPip(Pel)_0282 .....T.....LS..... 344
wPip(JHB)1 .....T.....LS..... 247
wPip(JHB)2 .....T.....LS..... 344
wPip(Pel)_0294 ---------.TT.....QV....DLSEETKKCM.NV..R--IIKLKTQDRGEQ----S 330
wPip(Buckeye) TSGVFAELIVNWRREDIYKTDEEKEIEKKEIL-DMMSFAKDCFPEKFELFKKLIIRDLRL 403
wPip(Pel)_0282 .....-..... 403
wPip(JHB)1 .....-..... 306
wPip(JHB)2 .....-..... 403
wPip(Pel)_0294 I.SI.VDYS.TYTIANLIVDPSRQGVS.E...GKILKHV.EMSG.EMIKV.DSVLSKIQ. 390
wPip(Buckeye) CGREGKRVNVDYGLFAEELFSELEKTILPPGPVGDGPCSNLRSRSKAHGSKKTTLPVDDS 463
wPip(Pel)_0282 463
wPip(JHB)1 366
wPip(JHB)2 463
wPip(Pel)_0294 FHG-..KLQLG-----.QV..K.AQEASKE--------.I..EAG------------.TL 424
wPip(Buckeye) PQSELGTPSVSGVSSYKKKSVFTLSGNK 491
wPip(Pel)_0282 491
wPip(JHB)1 394
wPip(JHB)2 491
wPip(Pel)_0294 ...S.S.TDTP-------YNIKS..HS. 445

Figure 2. Total mass spectrometry detected peptide coverage of WPIP0282 and a sequence alignment of homologs from *Culex pipiens* strains of *Wolbachia*. Dots indicate amino acid identities with respect to *w*Pip(Buckeye). Black shaded residues indicate peptides detected by mass spectrometry in both spermathecal and ovarian samples. Grey shaded residues indicate peptides detected only in ovarian samples. In *w*Pip (Pel)_0282 homologs, vertical boxes indicate regions of divergence with respect to *w*Pip(Buckeye). The bold black underline indicates a region of divergence in *w*Pip(JHB)2. *w*Pip (Pel)_0294 is a distant homolog of *w*Pip (Pel)_0282. Asterisks indicate fully conserved residues.

*w*Pip (JHB)2, located approximately 120 residues downstream of the N-terminus (underlined in Fig 2) and a single A/V substitution at residue 255. The three homologs clearly indicate that the *w*Pip (JHB) genome encodes at least two close homologs of *w*Pip (Pel)_0282. Moreover, as will be described in further detail below, BlastP analysis using *w*Pip (Pel)_0282 sequence as the subject query uncovered two accessions with ~ 34% identity in *w*AlbB (ZP_09542503.1 and ZP_09542120.1; see Fig 3), a strain of *Wolbachia* found in the Culicine mosquito, *Aedes albopictus*. Finally, as shown in the bottom entry of the alignment in Fig 2, BlastP uncovered a more divergent homolog, *w*Pip_0294 (YP_001975095) in the annotated *w*Pip (Pel) genome. With introduction of gaps, the 0294 protein had low homology to WPIP0282, with 132/410 (32%) conserved residues and 210/410 (51%) similarities.

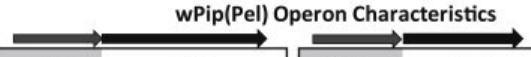
Subsequently, using the *w*Pip (Pel)_0294 homolog sequence as the BlastP subject query, we found that *w*AlbB protein (ZP_09542120.1) had 97% identity, but among JHB accessions, *w*Pip (Pel)_0294 had no match closer than *w*Pip_0282 (Fig 2). Overall these comparisons demonstrated that at least two homologs of the WPIP0282 protein exist within *Wolbachia* from both *C. pipiens* and *A. albopictus* mosquitoes. In all three bacterial genomes: *w*Pip (Pel), *w*Pip (JHB), and *w*AlbB, one of these variants retains close resemblance to the WPIP0282 sequence identified by mass spectrometry, while the other more closely resembles the divergent 0294-like variant in *w*Pip (Pel) and shares a pattern of consensus amino acids suggestive of duplication and divergence from an ancestral gene.

Homologs of WPIP0282 are Unique to CI-inducing Wolbachia. To evaluate whether WPIP0282 might participate in CI and to compile a list of homologs and their relationships, we used BlastP analysis to search for homologs more broadly among the *Rickettsiales*. Interestingly, we found that WPIP0282 homologs were unique to the genus *Wolbachia*, and absent from the *Rickettsia*, *Anaplasma*, and *Ehrlichia*. Furthermore, within the *Wolbachia*, homologs occur only in insect-associated, CI-inducing strains such as *w*Pip, *w*Ri, *w*Mel, *w*Sim, *w*VitB, *w*Ha, *w*No, and *w*AlbB; WPIP0282 homologs were consistently absent from the *Wolbachia* genomes of non-CI-inducing, obligate mutualistic nematode strains *w*Bm and *w*Oo.

Among *Wolbachia* that infect insects, genomes have been fully annotated for *w*Pip (Pel), and from two strains that infect *Drosophila*: *w*Mel (*Drosophila melanogaster*) and *w*Ri (Riverside strain of *Drosophila simulans*). As summarized in Fig 3, the *w*Ri genome encodes three well-conserved homologs of *w*Pip_0282 while the *w*Mel genome encodes only one, *w*Mel_0631. Alignment of WPIP0282 with those homologs from *Drosophila* associated *Wolbachia* strains (Fig 4) showed an overall conservation of 36% identities (179/491 residues) among all homologs, and sites that differed from WPIP0282 in *w*Mel or *w*Ri are identical in the two *Drosophila* strains, with the exception of greater divergence in *w*Ri_006720. The longest region of conservation occurs in the center of the protein, whereas the N-terminal region, including the ~ 20 residue divergent span noted previously for WPIP0282 proteins in *C. pipiens* strains (underlined in Fig 2 and Fig 4) may be evolving more quickly and/or may be less important for protein function. Near the C-terminus, the homologs of *Wolbachia*

A

wPip(Pel) Operon Characteristics



wPip(Pel)	0282 GI:190570728	0283 GI:190570729	0294 GI:190570737	0295 GI:190570738	1291 GI:190571665	1292 GI:190571666
Amino Acid residues	491	1174	445	732	353	118
Domains		FAD-binding, phosphate-binding peptidase C48; pfam02902		DUF-1703	peptidase C48; pfam02902	
Homology to 0282/0283	to WPIP0282	to WPIP0283	to WPIP0282	to WPIP0283	to WPIP0283 (C-term)	to WPIP0283 (C-term)
(identity%/similarity%)	100/100	100/100	34/53	23/42	99/99	38/54
E Value	0.0	0.0	2e-44	1e-24	0.0	4e-05

B

Homolog BLASTp Comparisons

Homology to:	0282	0283	0294	0295	1291	1292
wMel Operon	*w* Mel_0631	*w* Mel_0632	*w* Mel_0631	*w* Mel_0632	*w* Mel_0632	*w* Mel_0632
residue overlap	488	1180	376	552	353	108
identity%/similarity%	66/77	76/83	30/51	23/42	87/92	45/56
E Value	0.0	0.0	4e-41	6e-19	0.0	6e-07
wRi Operon	*w* Ri_005370	*w* Ri_p05380	*w* Ri_005370	*w* Ri_p05380	*w* Ri_p05380	*w* Ri_p05380
residue overlap	488	950	376	557	353	108
identity%/similarity%	66/77	74/82	30/51	22/42	86/92	45/56
E Value	0.0	0.0	1e-41	1e-15	0.0	6e-08
wRi Operon	*w* Ri_010030	wRi_p010040	wRi_010030	wRi_p010040	wRi_p010040	wRi_p010040
residue overlap	488	950	376	557	353	108
identity%/similarity%	66/77	74/82	30/51	22/42	86/92	45/56
E Value	0.0	0.0	1e-41	1e-15	0.0	6e-08
wRi Operon	wRi_006720	wRi_006710	wRi_006720	wRi_006710	none	none
residue overlap	495	802	460	729		
identity%/similarity%	44/60	30/45	30/51	31/51		
E Value	8e-120	9e-61	1e-47	4e-78		
wAlbB Operon	ZP_09542503	ZP_09542504	ZP_09542503	ZP_09542504	none	none
residue overlap	402	778	375	727		
identity%/similarity%	35/49	30/47	28/45	31/51		
E Value	4e-45	7e-63	9e-23	9e-73		
wAlbB Operon	ZP_09542120	ZP_09542121	ZP_09542120	ZP_09542121	none	none
residue overlap	440	784	445	732		
identity%/similarity%	32/50	22/42	97/98	98/98		
E Value	4e-43	3e-26	0.0	0.0		

Figure 3. Tabled array of various protein homologs analyzed in this study. **A.** Organization, characteristics, and BlastP comparisons of all the proteins encoded by the genome of *w*Pip (Pel): WPIP0282/0283, WPIP0294/0295, WPIP1291 and WPIP1292. Arrows at top indicate organization of genes and direction of transcription. Grey-shaded arrows and boxes indicate homologs of *w*Pip_0282; black arrows and unshaded text indicate homologs of *w*Pip_0283. Identity, similarity, and E values are based on amino acid residues. **B.** BlastP comparisons of homologs in various *Wolbachia* strains queried against homologs in *w*Pip (Pel).

associated with *Drosophila* have a strong region of positive charge (HKKRK) partially conserved in the *Culex* homologs, with the exception of the more divergent protein encoded by *w*Pip_0294 (compare Figs 2 and 4). These analyses further verify the presence of multiple versions of the WPIP0282 protein in at least one *Wolbachia* strain in *Drosophila* (*w*Ri) and again suggest the gene has undergone distinct patterns of divergence among the *Wolbachia* lineages associated with *Drosophila* and *Culicines*.

wPip_0282 is Part of a Conserved, Two Gene Operon that has Undergone Gene Duplication. On the *w*Pip (Pel) chromosome, *w*Pip_0282 and its more distant relative, *w*Pip_0294, are separated by only a few genes. To evaluate whether these homologs diverged following a gene duplication event, we compared their immediate genomic environment in the available annotated *Wolbachia* genomes, and in *Wolbachia* accessions from the mosquito *Aedes albopictus* (Fig 3). Both *w*Pip_0282 and *w*Pip_0294 are each members of two gene operons: *w*Pip_0282 is immediately upstream of *w*Pip_0283, which encodes a 1174 residue protein, with a predicted mass of 134 kDa, and pI of 5.83. *w*Pip_0282/0283 genes are oriented in the same direction, and are separated by only 54 nucleotides. Moreover, the closest homolog (Blast) to *w*Pip_0283 is *w*Mel_0632, which lies immediately downstream of the *w*Pip_0282 homolog encoded by *w*Mel_0631 (Fig 3). Thus, in the genomes of *w*Pip (Pel) and *w*Mel, the 0282/0283 and 0631/0632 gene pairs are homologous and arranged in the same order. Similarly, *w*Pip_0282 homologs in *w*Ri (*w*Ri_005370, *w*Ri_006720, and *w*Ri_010030) are each followed by a homolog of *w*Pip_0283. Moreover, homologous protein accessions (BlastP) from the un-annotated genomes of *w*Pip (JHB) (accessions ZP_03335574 and ZP_03335382; not shown) and

wAlbB (ZP_09542121, ZP09542504; Fig 3) further support broad conservation of the 0282/0283 gene pair. Likewise, the distant homolog of *w*Pip_0282 encoded by *w*Pip_0294 also appears to be part of a two-gene operon, separated by only 7 nucleotides from *w*Pip_0295 (Fig 3), which is distantly related (BlastP) to *w*Pip_0283 (Fig 3) with introduction of gaps. Thus, in the genome of *w*Pip (Pel), the 0282/0283 and 0294/0295 gene pairs are homologous. However, *w*Pip_0295 entirely lacks 384 residues at the C-terminus (Fig 5). Alignments of the second gene in the operons within the genome of *w*Pip (Pel) are shown in Fig 5.

Adding even more complexity is the fact that the 384-residue C terminus of the protein encoded by *w*Pip (Pel)_0283 (and absent from the *w*Pip_0295 homolog) is represented in its entirety by *w*Pip_1291, with a single L/M replacement at its N-terminus (see the black circle followed by a blackened entry in Figure 5; white dots on a black background indicate amino acid identity). Both WPIP0283 and WPIP1291 share in common a 90 amino acid motif that corresponds to a sentrin/SUMO specific protease domain in the Ulp1 protease family (pfam02902) (Fig 5, denoted by asterisks) as determined by NCBI's Conserved Domain Database and EMBL-EBI's InterProScan (Zdobnov and Apweiler, 2001) and maintain conservation of the essential catalytic histidine, aspartic acid, glutamine, and cysteine residues identified by Li and Hochstrasser (1999) and by Mossessova and Lima (2000) (Fig 5 upward black arrows). Finally, downstream of this conserved domain, the adjacent *w*Pip_1292 encodes a 118 amino acid sequence with homology to the extreme C-terminus encoded by *w*Pip (Pel)_0283 (Fig 5, see the entry boxed in grey and preceded by a grey, filled circle).

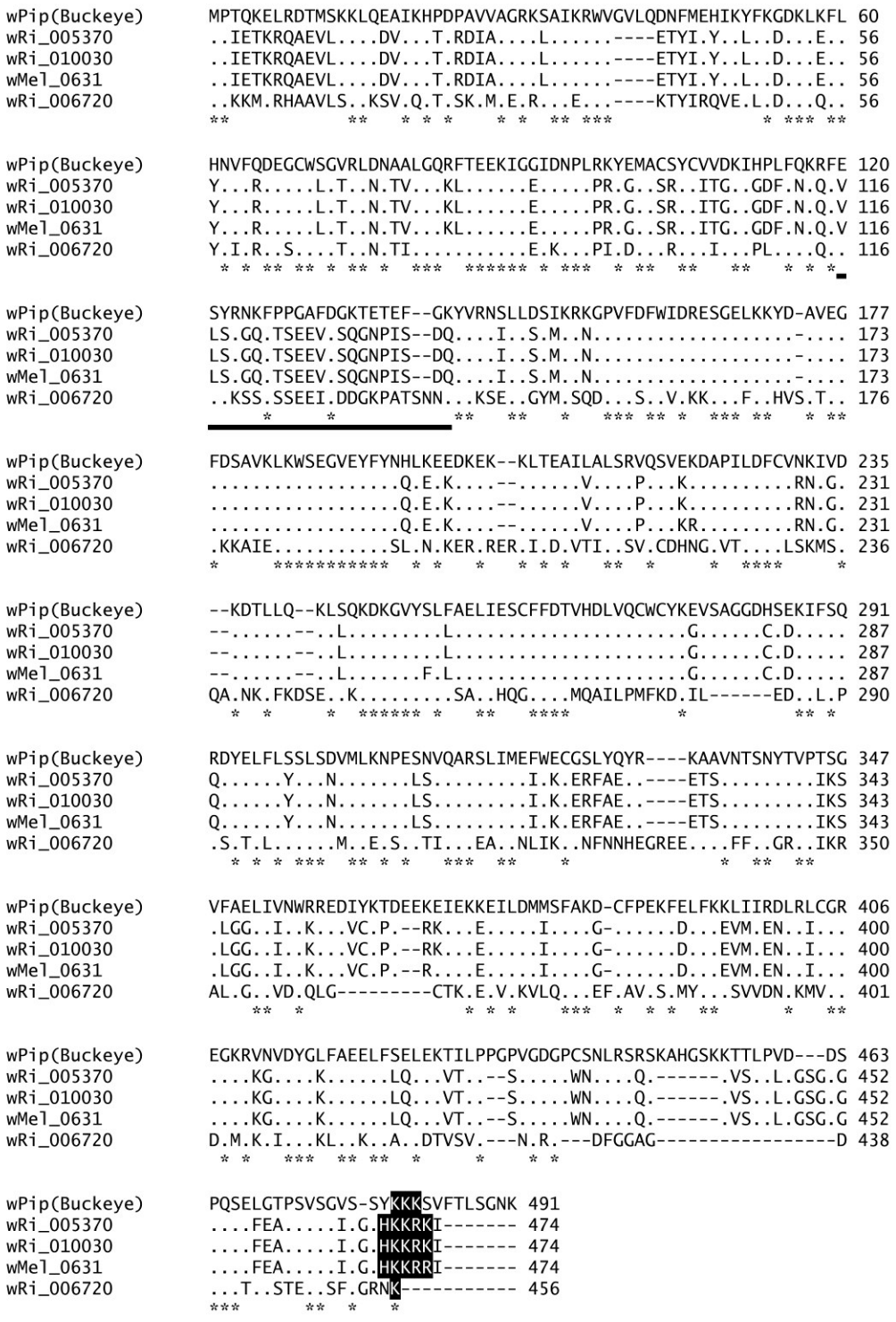

wPip(Buckeye) MPTQKELRDTMSKKLQEAIKHPDPAVVAGRKSAIKRWVGVLQDNFMEHIKYFKGDKLKFL 60
wRi_005370 ..IETKRQAEVL....DV...T.RDIA....L......----ETYI.Y..L..D...E.. 56
wRi_010030 ..IETKRQAEVL....DV...T.RDIA....L......----ETYI.Y..L..D...E.. 56
wMel_0631 ..IETKRQAEVL....DV...T.RDIA....L......----ETYI.Y..L..D...E.. 56
wRi_006720 ..KKM.RHAAVLS..KSV.Q.T.SK.M.E.R...E...----KTYIRQVE.L.D...Q.. 56
wPip(Buckeye) HNVFQDEGCWSGVRLDNAALGQRFTEEKIGGIDNPLRKYEMACSYCVVDKIHPLFQKRFE 120
wRi_005370 Y...R.....L.T..N.TV...KL......E.....PR.G..SR..ITG..GDF.N.Q.V 116
wRi_010030 Y...R.....L.T..N.TV...KL......E.....PR.G..SR..ITG..GDF.N.Q.V 116
wMel_0631 Y...R.....L.T..N.TV...KL......E.....PR.G..SR..ITG..GDF.N.Q.V 116
wRi_006720 Y.I.R..S....T..N.TI..........E.K...PI.D...R...I...PL....Q.. 116
wPip(Buckeye) SYRNKFPPGAFDGKTETEF--GKYVRNSLLDSIKRKGPVFDFWIDRESGELKKYD-AVEG 177
wRi_005370 LS.GQ.TSEEV.SQGNPIS--DQ....I..S.M..N.....-.... 173
wRi_010030 LS.GQ.TSEEV.SQGNPIS--DQ....I..S.M..N.....-.... 173
wMel_0631 LS.GQ.TSEEV.SQGNPIS--DQ....I..S.M..N.....-.... 173
wRi_006720 ..KSS.SSEEI.DDGKPATSNN...KSE..GYM.SQD...S..V.KK...F..HVS.T.. 176
wPip(Buckeye) FDSAVKLKWSEGVEYFYNHLKEEDKEK--KLTEAILALSRVQSVEKDAPILDFCVNKIVD 235
wRi_005370 .....Q.E.K....--.....V....P...K.....RN.G. 231
wRi_010030 .....Q.E.K....--.....V....P...K.....RN.G. 231
wMel_0631 .....Q.E.K....--.....V....P...KR.....RN.G. 231
wRi_006720 .KKAIE.....SL.N.KER.RER.I.D.VTI..SV.CDHNG.VT....LSKMS. 236
wPip(Buckeye) --KDTLLQ--KLSQKDKGVYSLFAELIESCFFDTVHDLVQCWCYKEVSAGGDHSEKIFSQ 291
wRi_005370 --.....--..L.....L.....G.....C.D..... 287
wRi_010030 --.....--..L.....L.....G.....C.D..... 287
wMel_0631 --.....--..L.....F.L.....G.....C.D..... 287
wRi_006720 QA.NK.FKDSE..K.....SA..HQG....MQAILPMFKD.IL------ED..L.P 290
wPip(Buckeye) RDYELFLSSLSDVMLKNPESNVQARSLIMEFWECGSLYQYR----KAAVNTSNYTVPTSG 347
wRi_005370 Q.....Y...N.....LS.....I.K.ERFAE..----ETS.....IKS 343
wRi_010030 Q.....Y...N.....LS.....I.K.ERFAE..----ETS.....IKS 343
wMel_0631 Q.....Y...N.....LS.....I.K.ERFAE..----ETS.....IKS 343
wRi_006720 .S.T.L.....M..E.S..TI...EA..NLIK..NFNNHEGREE....FF..GR..IKR 350
wPip(Buckeye) VFAELIVNWRREDIYKTDEEKEIEKKEILDMMSFAKD-CFPEKFELFKKLIIRDLRLCGR 406
wRi_005370 .LGG..I..K...VC.P.--RK...E.....I....G-.....D...EVM.EN..I... 400
wRi_010030 .LGG..I..K...VC.P.--RK...E.....I....G-.....D...EVM.EN..I... 400
wMel_0631 .LGG..I..K...VC.P.--R....E.....I....G-.....D...EVM.EN..I... 400
wRi_006720 AL.G..VD.QLG---------CTK.E.V.KVLQ...EF.AV.S.MY...SVVDN.KMV.. 401
wPip(Buckeye) EGKRVNVDYGLFAEELFSELEKTILPPGPVGDGPCSNLRSRSKAHGSKKTTLPVD---DS 463
wRi_005370 ....KG....K.....LQ...VT..--S.....WN....Q.------.VS..L.GSG.G 452
wRi_010030 ....KG....K.....LQ...VT..--S.....WN....Q.------.VS..L.GSG.G 452
wMel_0631 ....KG....K.....LQ...VT..--S.....WN....Q.------.VS..L.GSG.G 452
wRi_006720 D.M.K.I...KL..K..A..DTVSV.---N.R.---DFGGAG----------------D 438
wPip(Buckeye) PQSELGTPSVSGVS-SYKKKSVFTLSGNK 491
wRi_005370 ....FEA.....I.G.HKKRKI------- 474
wRi_010030 ....FEA.....I.G.HKKRKI------- 474
wMel_0631 ....FEA.....I.G.HKKRRI------- 474
wRi_006720 ...T..STE..SF.GRNK---------- 456

Figure 4. Protein alignments of other WPA0282 homologs within *w*Mel and *w*Ri. Dots indicate identities with respect to *w*Pip(Buckeye) and asterisks indicate fully conserved residues. The bold black underline highlights the region of divergence corresponding to that shown for *w*Pip(JHB)2 in Fig 2. Black shaded residues are conserved C-terminal positively charged residues.

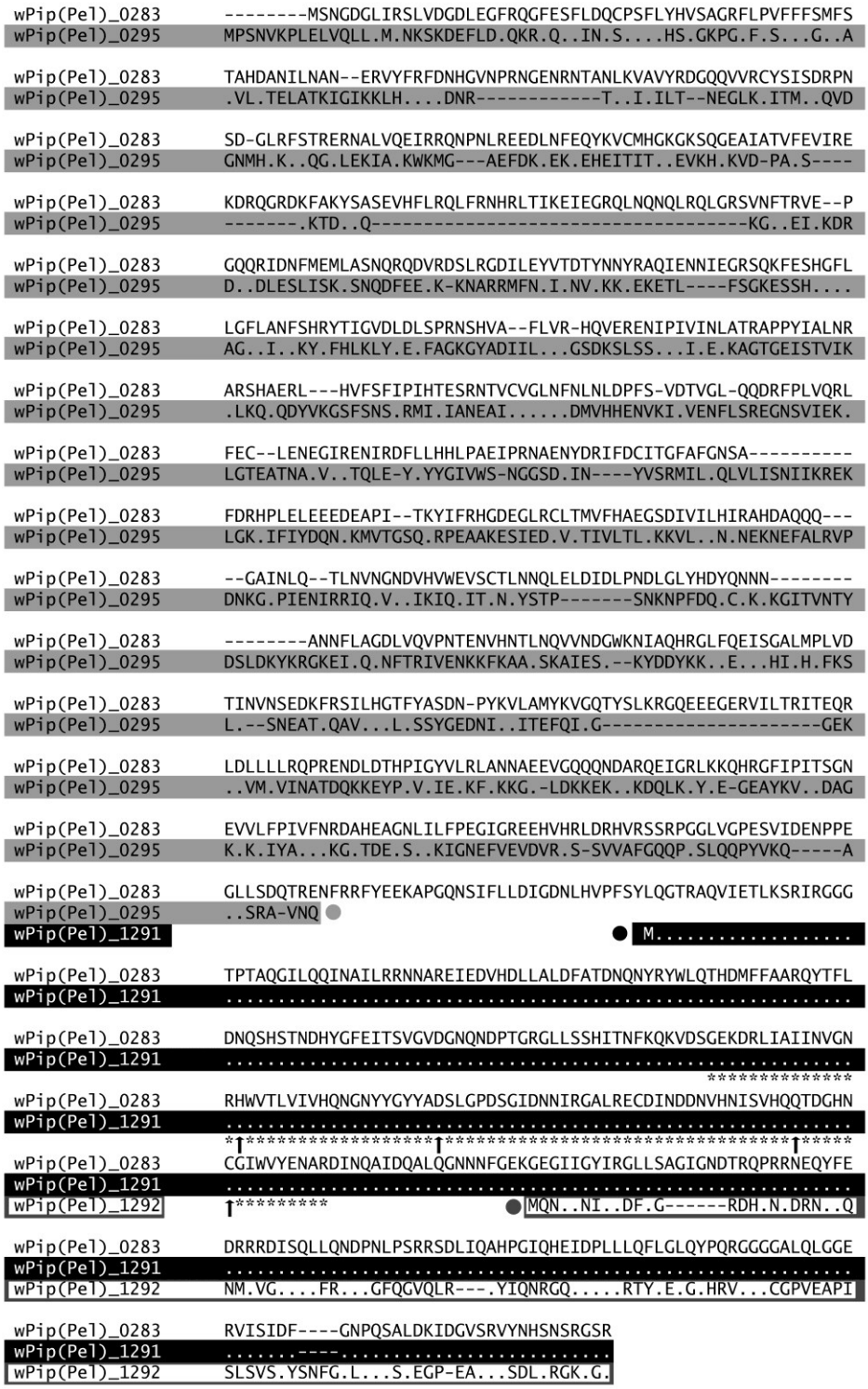

wPip(Pel)_0283 --------MSNGDGLIRSLVDGDLEGFRQGFESFLDQCPSFLYHVSAGRFLPVFFFSMFS
wPip(Pel)_0295 MPSNVKPLELVQLL.M.NKSKDEFLD.QKR.Q..IN.S....HS.GKPG.F.S...G..A
wPip(Pel)_0283 TAHDANILNAN--ERVYFRFDNHGVNPRNGENRNTANLKVAVYRDGQQVVRCYSISDRPN
wPip(Pel)_0295 .VL.TELATKIGIKKLH....DNR------------T..I.ILT--NEGLK.ITM..QVD
wPip(Pel)_0283 SD-GLRFSTRERNALVQEIRRQNPNLREEDLNFEQYKVCMHGKGKSQGEAIATVFEVIRE
wPip(Pel)_0295 GNMH.K..QG.LEKIA.KWKMG---AEFDK.EK.EHEITIT..EVKH.KVD-PA.S----
wPip(Pel)_0283 KDRQGRDKFAKYSASEVHFLRQLFRNHRLTIKEIEGRQLNQNQLRQLGRSVNFTRVE--P
wPip(Pel)_0295 -------.KTD..Q-----------------------------------KG..EI.KDR
wPip(Pel)_0283 GQQRIDNFMEMLASNQRQDVRDSLRGDILEYVTDTYNNYRAQIENNIEGRSQKFESHGFL
wPip(Pel)_0295 D..DLESLISK.SNQDFEE.K-KNARRMFN.I.NV.KK.EKETL----FSGKESSH....
wPip(Pel)_0283 LGFLANFSHRYTIGVDLDLSPRNSHVA--FLVR-HQVERENIPIVINLATRAPPYIALNR
wPip(Pel)_0295 AG..I..KY.FHLKLY.E.FAGKGYADIIL...GSDKSLSS...I.E.KAGTGEISTVIK
wPip(Pel)_0283 ARSHAERL---HVFSFIPIHTESRNTVCVGLNFNLNLDPFS-VDTVGL-QQDRFPLVQRL
wPip(Pel)_0295 .LKQ.QDYVKGSFSNS.RMI.IANEAI.....DMVHHENVKI.VENFLSREGNSVIEK.
wPip(Pel)_0283 FEC--LENEGIRENIRDFLLHHLPAEIPRNAENYDRIFDCITGFAFGNSA----------
wPip(Pel)_0295 LGTEATNA.V..TQLE-Y.YYGIVWS-NGGSD.IN----YVSRMIL.QLVLISNIIKREK
wPip(Pel)_0283 FDRHPLELEEEDEAPI--TKYIFRHGDEGLRCLTMVFHAEGSDIVILHIRAHDAQQQ---
wPip(Pel)_0295 LGK.IFIYDQN.KMVTGSQ.RPEAAKESIED.V.TIVLTL.KKVL..N.NEKNEFALRVP
wPip(Pel)_0283 --GAINLQ--TLNVNGNDVHVWEVSCTLNNQLELDIDLPNDLGLYHDYQNNN--------
wPip(Pel)_0295 DNKG.PIENIRRIQ.V..IKIQ.IT.N.YSTP-------SNKNPFDQ.C.K.KGITVNTY
wPip(Pel)_0283 --------ANNFLAGDLVQVPNTENVHNTLNQVVNDGWKNIAQHRGLFQEISGALMPLVD
wPip(Pel)_0295 DSLDKYKRGKEI.Q.NFTRIVENKKFKAA.SKAIES.--KYDDYKK..E...HI.H.FKS
wPip(Pel)_0283 TINVNSEDKFRSILHGTFYASDN-PYKVLAMYKVGQTYSLKRGQEEEGERVILTRITEQR
wPip(Pel)_0295 L.--SNEAT.QAV...L.SSYGEDNI..ITEFQI.G----------------------GEK
wPip(Pel)_0283 LDLLLLRQPRENDLDTHPIGYVLRLANNAEEVGQQQNDARQEIGRLKKQHRGFIPITSGN
wPip(Pel)_0295 ..VM.VINATDQKKEYP.V.IE.KF.KKG.-LDKKEK..KDQLK.Y.E-GEAYKV..DAG
wPip(Pel)_0283 EVVLFPIVFNRDAHEAGNLILFPEGIGREEHVHRLDRHVRSSRPGGLVGPESVIDENPPE
wPip(Pel)_0295 K.K.IYA...KG.TDE.S..KIGNEFVEVDVR.S-SVVAFGQQP.SLQQPYVKQ-----A
wPip(Pel)_0283 GLLSDQTRENFRRFYEEKAPGQNSIFLLDIGDNLHVPFSYLQGTRAQVIETLKSRIRGGG
wPip(Pel)_0295 ..SRA-VNQ
wPip(Pel)_1291 M.....
wPip(Pel)_0283 TPTAQGILQQINAILRRNNAREIEDVHDLLALDFATDNQNYRYWLQTHDMFFAARQYTFL
wPip(Pel)_1291 .....
wPip(Pel)_0283 DNQSHSTNDHYGFEITSVGVDGNQNDPTGRGLLSSHITNFKQKVDSGEKDRLIAIINVGN
wPip(Pel)_1291 .....
****
wPip(Pel)_0283 RHWVTLVIVHQNGNYYGYYADSLGPDSGIDNNIRGALRECDINDDNVHNISVHQQTDGHN
wPip(Pel)_1291 .....
wPip(Pel)_0283 CGIWVYENARDINQAIDQALQGNNNFGEKGEGIIGYIRGLLSAGIGNDTRQPRRNEQYFE
wPip(Pel)_1291 .....
wPip(Pel)_1292 MQN..NI..DF.G------RDH.N.DRN..Q
wPip(Pel)_0283 DRRRDISQLLQNDPNLPSRRSDLIQAHPGIQHEIDPLLLQFLGLQYPQRGGGGALQLGGE
wPip(Pel)_1291 .....
wPip(Pel)_1292 NM.VG....FR...GFQGVQLR---.YIQNRGQ.....RTY.E.G.HRV...CGPVEAPI
wPip(Pel)_0283 RVISIDF----GNPQSALDKIDGVSRVYNHSNSRGSR
wPip(Pel)_1291 .....----.....
wPip(Pel)_1292 SLSVS.YSNFG.L...S.EGP-EA...SDL.RGK.G.

Figure 5. Protein alignments of the second gene in the operon, *w*Pip0283, and its homologs in *w*Pip(Pel). Dots indicate identities with respect to *w*Pip (Pel)_0283. Grey shaded residues show the C-terminally truncated homolog, *w*Pip0295; its terminus is indicated by a grey circle. Black shaded residues show N-terminally truncated homolog, *w*Pip1291; its N-terminus is indicated by a black circle. At the bottom of the alignment dark grey boxed residues show the second N-terminally truncated homolog, *w*Pip1292; its N-terminus is shown by the dark grey circle. Asterisks indicate the eukaryotic Ulp1 Ubiquitin-like C48 SUMO protease domain (pfam02902) with conserved catalytic residues identified by upward black arrows (Li and Hochstrasser, 1999; Mossessova and Lima, 2000). We note that the N-terminus of *w*Pip_1291 begins downstream of the C-terminus of *w*Pip_0295.

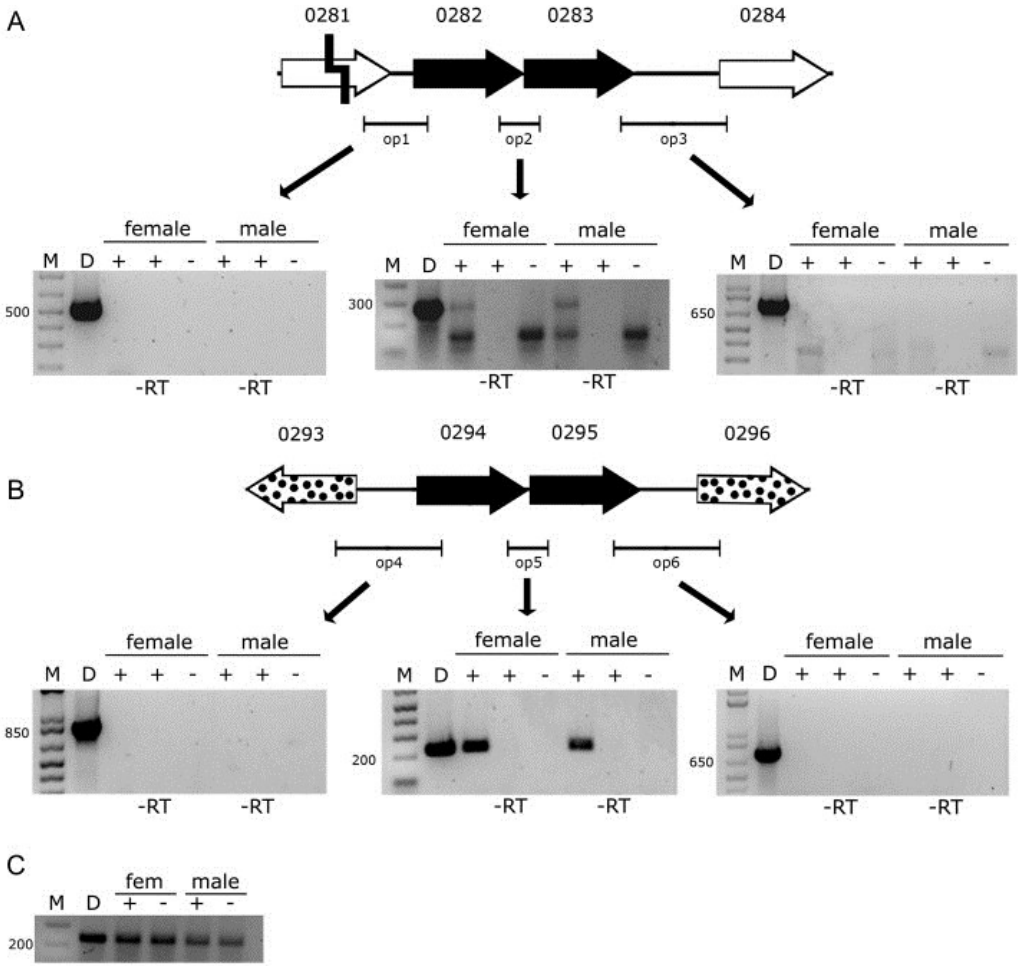

A
0281
0282
0283
0284
op1
op2
op3
female
male
M D + + - + + -
500
300
650
-RT
-RT
B
0293
0294
0295
0296
op4
op5
op6
850
200
650
C
fem
male
M D + - + -
200

Fig 6. RT-PCR confirmation of operon structure and expression within *Wolbachia* infected female and male *Culex pipiens* mosquitoes. **A.** RT-PCR analysis of 0282-0283 operon. M is a DNA marker. D is a *Wolbachia* DNA positive PCR control. At the top of the gels, Plus (+) and minus (-) signify *Wolbachia* infection status. Negative controls were performed using uninfected (-) female and male mosquitoes; and lanes labeled –RT on bottom signify a reaction without reverse transcriptase as a control for DNA contamination of RNA samples. Positions of primers are indicated by horizontal bars labeled op1, op2, and op3. Bands were excised, sequenced, and determined to be the correct PCR/RT-PCR product. **B.** RT-PCR analysis of 0294-0295 operon. Symbols are as in A. Bands were excised, sequenced, and determined to be the correct PCR/RT-PCR product. **C.** Quality of RNA samples assayed by RT-PCR using primers for the mosquito ribosomal protein RpS3. Primer attributes are listed in table S1.

In summary, there is one full (*w*Pip_0283) and three partial copies of the second operon gene (*w*Pip_0295, *w*Pip_1291, and *w*Pip_1292) within the genome of *w*Pip (Pel) and two of those genes (*w*Pip_0283 and *w*Pip_1291) have SUMO protease domains.

Gene pairs 0282/0283 and 0294/0295 are expressed as operons in Wolbachia-infected mosquitoes. We hypothesized that the wPip (Pel)_0282/0283 gene pair and their more distant homologs 0294/0295 are organized as operons because they are adjacent, oriented in the same direction, and are separated by fewer than 60 nucleotides. We used Reverse Transcriptase (RT)-PCR to verify expression of WPIP0282 at the mRNA level, and to explore whether 0282 and 0283 are transcribed as one polycistronic mRNA molecule. Primers that spanned the intergenic regions between genes 0282/0283, produced the predicted products, while products corresponding to flanking genes (pseudogene 0281 and distal 0284) were not produced (Fig 6A). We detected expression of the 0282/0283 operon in both male and female *Culex pipiens* Buckeye mosquitoes, indicating transcription in testes and ovaries. We note that another 150 base pair RT-PCR band was produced with the op2 primers. This band was likely a product of primers binding to another RNA sequence in the mosquitoes as the band appeared in both infected and uninfected host samples.

Similarly *w*Pip (Pel)_0294/0295, were also polycistronic, and transcribed separately from flanking, *w*Pip (Pel)_ 0293 and 0296 genes (Fig 6B). Again, we detected expression of the 0294-0295 polycistronic mRNA in both male and female mosquitoes,

indicating that expression occurs in testes and ovaries. RNA quality was verified by amplifying RNA extracts with primers for the mosquito ribosomal protein RpS3 (Fig 6C).

Expression in cultured cells. After we had defined the 0282/0283 and 0294/0295 operons we sought to determine if they were expressed in a tissue specific manner or also in cell culture. Unfortunately, *w*Pip has not been cultivated in cell culture, whereas *w*AlbB has. As shown in Fig 3, *w*AlbB homologs of WPIP0282: ZP_09542503.1 which is closer (BlastP) to WPIP0282 and ZP_09542120.1 which is more similar to WPIP0294. We designed primers for both *w*AlbB homologs and performed RT-PCR analysis in both TW-280 cells infected with *w*AlbB and a colony of *Aedes albopictus* mosquitoes dually infected with *w*AlbA and *w*AlbB. Both the 0282 and 0294 homologs were expressed in both cell culture and within the male and female mosquitoes, suggesting that the *w*AlbB 0282 and 0294 homologs are not expressed in a gonadal specific manner (data not shown).

***Genomic context of wPip_0282/0283 homologs*.** Finally, although their relative positions vary in available annotated *Wolbachia* genomes (Fig 7A), the 0282/0283 gene pairs in *w*Pip (Pel) and their homologs in *w*Ri (summarized in Fig 3) are immediately surrounded by mobile genetic elements (MGE's) that potentially encode integrase (I)/recombinase (RC)/transposase (T) and/or reverse transcriptase (RT) proteins (Fig 7B, open arrows), all of which are embedded near a cluster of WO-phage genes. The distal *w*Pip 0294/0295 pair is immediately surrounded by ankyrin repeat genes, which are thought to have been acquired from a eukaryotic host and potentially interact with host

proteins (Siozios et al., 2013). In the genomes of *w*Mel and *w*Ri, we note similar associations with genes that encode phage proteins and ankyrin repeats. Curiously, in *w*Mel, there are fewer MGE's in close proximity to the single copy operon, but in *w*Pip and *w*Ri, where MGE's are present in abundance and flank the operon, there are two and three copies of the operon respectively. Of particular interest, with respect to *w*Pip_0283, are the pseudogenes *w*Ri_p005380 (nucleotides 573202 to 576723 in NC_012416.1) and *w*Ri_p010040 (nucleotides 1079661 to 1083182; Fig 7B, indicated by broken lines in the *w*Ri accessions). Curiously, Blast comparisons of nucleotide sequences indicate complete identity over 3522 residues, which is unusual if these are unexpressed pseudogenes. To investigate these accessions more closely, we translated nucleotides 573202 to 576723 in NC_012416.1. One of the reading frames contains a single stop codon interrupting the putative translation product encoded by *w*Ri_p005380. Upstream of the stop codon is an open reading frame encoding 105 amino acid residues (including the FAD-binding, phosphate-binding motif indicated in Fig 3), followed by a 29 codon, untranslated region, and a second, in-frame Methionine, that begins an open reading frame encoding 950 residues and a fully conserved Ulp1 C48 SUMO protease domain. Based on these characteristics, we hypothesize that these putative pseudogenes are in fact expressed.

Discussion:

Wolbachia pipientis 0282 as a CI Effector Candidate. We initiated these studies to explore whether the *Wolbachia*-encoded DNA binding protein, HU beta (Beckmann et al., 2013), was enriched in mature sperm from *Wolbachia* infected male *Culex pipiens*. To do this we dissected spermathecae from females mated with infected males and subjected protein extracts to SDS PAGE and mass spectrometry. We did not detect the presence of the HU beta in the spermathecal extracts, but we did detect peptides from a 56 kDa hypothetical protein encoded by *w*Pip_0282. Searches using available protein domain identifiers, including NCBI's conserved domain database and EMBL-EBI's InterProScan did not identify conserved domains or other features that would suggest a function for WPIP0282. However, pDomTHREADER (Lobley et al. 2009) predicted that both WPIP0282 and WPIP0294 proteins show folding similarities resembling the crystal structure of Cullin-1. Cullin proteins are subunits of the large SCF (Skp, Cullin, F-box containing) ubiquitin ligase complex that interacts with numerous substrates within the cell. One particularly interesting homolog of Cullin, Apc2, is part of the anaphase-promoting complex, which functions in chromatid segregation during the metaphase/anaphase transition by ubiquitin-mediated proteolysis (Deshaies 1999). We also note that the predicted secondary structure of WPIP0282 includes multiple helical domains, and that the sequence contains 77 basic, as well as 77 acidic amino acid residues, with a predicted pI near neutrality.

To date, WPIP0282 is the only *Wolbachia*-encoded protein to be detected in mosquito sperm-filled spermathecae. Because *Wolbachia* is not known to reside in or

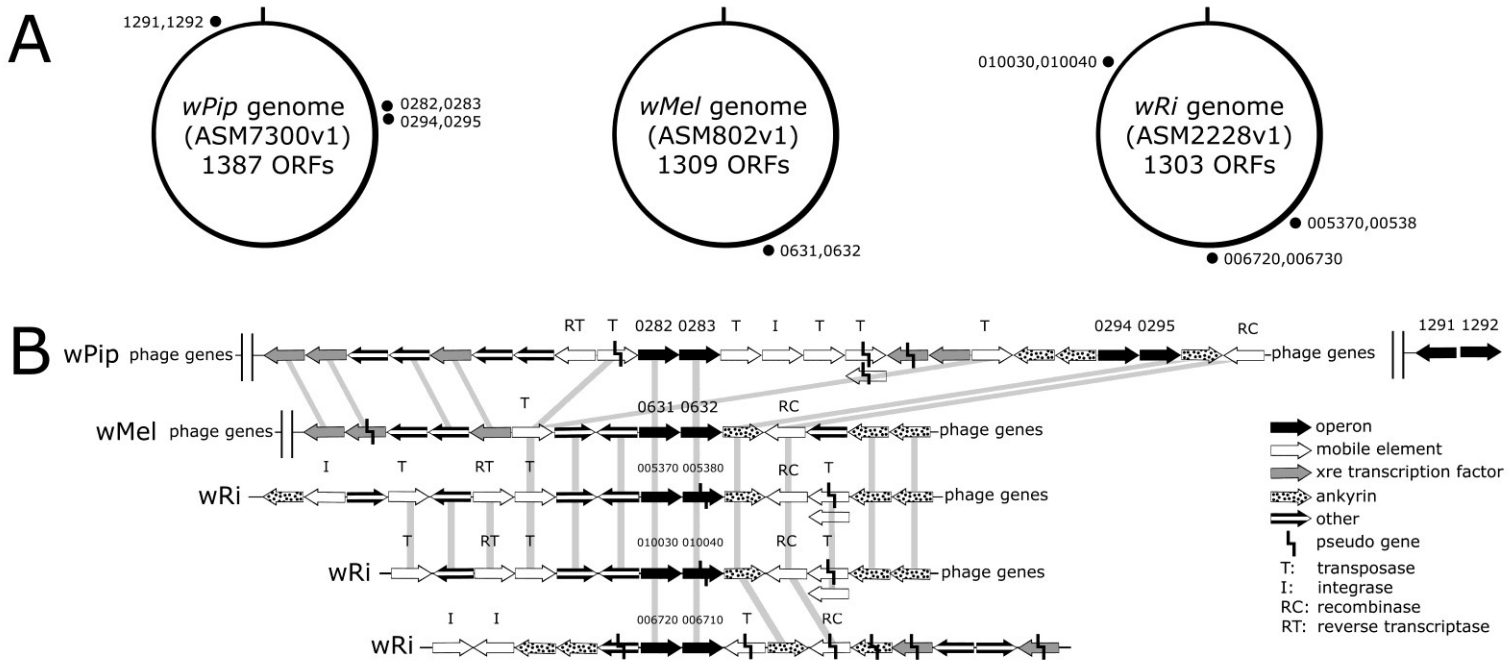

A
1291,1292
wPip genome
(ASM7300v1)
1387 ORFs
0282,0283
0294,0295
wMel genome
(ASM802v1)
1309 ORFs
0631,0632
010030,010040
wRi genome
(ASM2228v1)
1303 ORFs
005370,00538
006720,006730
B
wPip phage genes
wMel phage genes
wRi
wRi
wRi
phage genes
RT T 0282 0283 T I T T T 0294 0295 RC
1291 1292
0631 0632
005370 005380
010030 010040
006720 006710
operon
mobile element
xre transcription factor
ankyrin
other
pseudo gene
T: transposase
I: integrase
RC: recombinase
RT: reverse transcriptase

Figure 7. Operon structure and synteny within the three *Wolbachia* strains, *w*Pip endosymbiont of *Culex quinquefaciatus*, *w*Mel endosymbiont of *Drosophila melanogaster,* and *w*Ri endosymbiont of *Drosophila simulans*. **A.** Genomes of the three strains with locations of the duplicated operon elements labeled as black dots. In *w*Pip, the operon has been fully duplicated one time with parts of 0283 also being duplicated another two times to 1291 and 1292. In *w*Mel the operon only occurs once. In *w*Ri the operon has been duplicated three times throughout the genome. **B.** Expanded genome regions including the respective operon elements in the various species. Genes *w*Ri_005380 and *w*Ri_010040 are labeled as pseudo genes within the genome, but still encode a large open reading frame. Genes *w*Pip_0283, *w*Pip _1291, *w*Mel_0632, *w*Ri_p005380, and *w*Ri_p010040 all have a conserved eukaryotic C48 sumo protease protein domain. Gene sizes are not to scale. Vertical shading indicates synteny among genome regions.

near the spermathecae its presence among spermathecal proteins strongly suggests that WPIP0282 was secreted by *Wolbachia* in the testes and retained in mature mosquito sperm after elimination of *Wolbachia* in "waste bags" during spermatid elongation. Its association with sperm is compatible with a potential role in CI induction, and the gene appears to have been eliminated from the non-CI-inducing *Wolbachia* strains that are mutualistic endosymbionts in filarial nematodes. In infected insects, a range of CI potency has been experimentally observed (Bourtzis et al. 1996). We note that the copy number of 0282 homologs is higher in *w*Ri than in *w*Mel, and that *w*Ri causes a more severe CI phenotype than does *w*Mel within their respective natural hosts (Clark et al. 2003). Despite evidence for variation in gene copy number and sequence divergence, the potential role of WPIP0282 and/or WPIP0283 warrants further investigation. We are particularly interested in learning whether these proteins interact with the HU beta protein described previously (Beckmann et al. 2013).

Mechanisms of Mobility. Our in silico analysis of the 0282/0283 operon clearly demonstrated genome proximity to MGE's such as reverse transcriptases, transposases, integrases, recombinases, and resolvases as well as with genes encoding WO phage proteins. Because *Wolbachia* has a high percentage of MGE's, especially transposases (Bordenstein and Reznikoff 2005), it brings into question whether the association is simply coincidental or of some functional consequence. While many other genes are near mobile elements, gene duplication events in the reduced genomes of *Wolbachia* seem to be relatively rare. We envision that this operon is unique and its sequence divergence and duplications within *Wolbachia* genomes may have significant physiological

consequences and associated evolutionary roles. Such a hypothesis could explain why there are so many different incompatible species of *Culex pipiens* mosquitoes, all of which seem to harbor taxonomically similar *Wolbachia* strains (Klassen et al., 2008; Salzberg et al., 2008) yet are unable to produce viable offspring when crossed (Laven 1967, Magnin et al. 1987).

How the operon is duplicated is not clear. Mobile elements that "copy and paste," usually do so by means of an RNA intermediate (Benjamin et al., 2007). These elements are usually bounded by long terminal repeats which do not appear to be present near the operon. However, a group II intron-like reverse transcriptase gene, *w*Pip_0280, and an integrase gene, *w*Pip_0285, are both present near the operon, which makes it feasible that the entire operon might be contained within a large class I retro-element. An alternative theory is that operon duplications might simply be a by-product of the activity of adjacent reverse transcriptase genes and transposases, such as *w*Pip_0280 and *w*Pip_0291. The genes might also have been duplicated by virtue of their association with the WO phage; phages can act as shuttle vectors for other mobile elements (Bordenstein and Reznikoff, 2005). All three *Wolbachia* strains have operons bordering the edges of a WO phage (Fig 7). Interestingly, Ishmael et al. 2009 specifically highlighted the regions immediately surrounding the operon and near the WO phage as regions characterized by, "Rampant lateral phage transfer between diverse strains of *Wolbachia.*"

Insights from Duplicate Operons and Extrapolations to CI Functionality. We demonstrated by RT-PCR that *w*Pip_0282 is transcribed with *w*Pip_0283 as a polycistronic mRNA, as are the more divergent *w*Pip_0294/0295 pair. Although in-

silico analysis of the amino acid sequences of WPIP0282 and WPIP0294 did not provide insight as to the function of the operon, analysis of the second gene in each operon provided interesting clues to potential functionalities. WPIP0283 contains a eukaryotic Ulp1 C48 SUMO protease domain (pfam02902) which was likely acquired by a horizontal transfer from a eukaryotic host because bacteria are not known to use SUMOylation, except in cases where intracellular pathogens must interact with host protein machinery (Wimmer et al., 2012). Specifically, in *Yersinia*, another Ulp1 SUMO protease-like protein, YopJ, has been shown to play a regulatory role in the host cell's MAPK signaling pathway by decreasing levels of SUMO-1-conjugated proteins (Orth et al., 2000; Cornelis and Denecker, 2001). Ulp1 SUMO proteases were originally described in yeast, where it was shown that alterations in their functionality can lead to G2/M arrest and cell lethality (Li and Hochstrasser, 1999). Interestingly, Ulp1 C48 SUMO protease proteins are also commonly associated with architectural chromatin proteins such as SMC, (Structural Maintenance of Chromosome) proteins and histones which themselves become SUMOylated (Lee et al., 2011; McAleenan et al., 2012; Cubenas-Potts and Matunis, 2013). Alterations in the SUMOylation status of host SMC proteins and/or histones might contribute to induction/rescue of known CI chromatin disruption phenotypes. If the Ulp1 SUMO protease domain was indeed horizontally transferred the most likely donor candidates were flies, as the closest eukaryotic similarities outside of *Wolbachia* for the C-terminus of WPIP0283 occur in both *Drosophila virilis* (gi|195393912) and *Drosophila willistoni* (gi|195448669) homologs. Horizontally transfered DNA encoding this protein domain also strengthens the

hypothesis that the gene participates with MGE's, which can import and retain host protein domains when beneficial. In *w*Pip portions of 0283 have been duplicated in the genes *w*Pip_1291, and 1292; the former of which contains an identical copy of the Ulp1 C48 SUMO protease domain harbored in *w*Pip_0283. In *w*Ri, the two close homologs of this gene are labeled as pseudogenes (*w*Ri_p005380 and *w*Ri_p010040), but they maintain an extremely large open reading frame (950 residues) with a fully conserved and recognizable Ulp1 C48 SUMO protease domain; if these genes were truly pseudogenes it would be unlikely that they would maintain such large, fully conserved open reading frames. If *w*Pip_0283 is involved with CI, variations in its copy number and sequence may be responsible for not only variations in CI potency, but also for strains of *Wolbachia* that can rescue CI but not induce it. A deletion in *w*Pip_0282, that left *w*Pip_0283 intact could disrupt the operon and conceivably create such a strain of *Wolbachia* if WPIP0283 was a rescue factor.

WPIP0283 is not the only protein that includes an interesting functional domain. Its partially duplicated homolog, *w*Pip_0295, while not having a SUMO protease domain, encodes a DUF 1703 protein domain that is also present in another selfish genetic element, named the "Medea element", which is associated with very similar CI like phenotypes and embryonic death in *Tribolium castaneum* (Lorenzen et al. 2008). The DUF 1703 domain is likely to be a nuclease within the PD-(D/E)XK family (Knizewski et al. 2007) and the Medea element is part of a large Tc1 mariner transposon, which reinforces similarities between Medea and the *Wolbachia* operon.

If the 0282-0283 operon is involved in CI it logically gives rise to the hypothesis of CI induction/rescue as the result of a two gene operon where one gene induces and another rescues CI. This hypothesis is supported by the structural similarity of WPIP0282 to Cullin-1 which aides in ligating ubiquitin substrates, whereas the WPIP0283 protein contains the ubiquitin-like SUMO protease which may counteract the function of the former protein by a precisely opposite mechanism. A system such as this would be strikingly similar to toxic-antidote operons or "addiction genes" which are commonly incorporated into mobile elements to increase segregation persistence (Rankin et al. 2011). Like CI, such systems are in fact often linked to chromatin functionality; many well studied toxic antidote operons target enzymes such as DNA gyrase and helicase genes which exhibit their effects on chromatin (Yamaguchi et al. 2011). If a CI inducing/rescue system was as simple as a two gene operon it would have profound impacts on the application of CI for genetic control of insects. One would no longer have to struggle with infecting mosquitoes with *Wolbachia* for the purposes of reproductive manipulation. The operon could potentially be cloned under a mosquito gonad-specific promoter to create a CI inducing strain of transformed mosquito and accomplish gene drive simply by the activities of the operon. CI induction/rescue by a two gene operon in a MGE also brings forth new hypotheses as to the evolution and origin of CI. Like the "Medea element," CI might be a remnant of a bacterial selfish genetic element whose gene product(s) interact with chromosome maintenance machinery encoded by the host genome of the *Wolbachia*-infected arthropod. Further work must be done to elucidate the operon's possible role in CI, but this work highlights the most intriguing CI-inducing

protein candidates from *Wolbachia* thus far. We have cloned the gene and are working on these studies currently.

Acknowledgements:

We thank Todd Markowski and Bruce Witthuhn at the *University of Minnesota's Center for Mass Spectrometry and Proteomics* for help with mass spectrometry detection and gel staining, Cassie Kurtz and Elissa Carroll for help with mosquito maintenance, and Gerald Baldridge, Rod Felsheim, and Tim Kurtti for helpful discussions. This work was supported by NIH grant AI081322 and by the University of Minnesota Agricultural Experiment Station, St. Paul, MN.

Additional Unpublished Proteomic Results:

The Wolbachia Ovarian Proteome. *Wolbachia* infections occur primarily within arthropod ovaries wherein the *Wolbachia* are maternally transmitted and in some cases possibly provide nutritional supplementation to eggs as well as manipulate reproduction and fecundity. In order to better understand these interactions as well as supplement known cellular biology with respect to the *Wolbachia* lifestyle I performed comparative iTRAQ analysis of *Wolbachia* infected/uninfected ovaries. iTRAQ, isobaric tags for relative and absolute quantification, is a quantitative mass spectrometry analysis used to calculate differences in peptide and protein abundance. The purpose of this experiment was to not only compile a list proteins which are quantitatively up or down-regulated by

Wolbachia infection, but also to produce a *Wolbachia* infected ovary proteome. Finally we sought to compare this proteome with previous data that we had collected with respect to protein expression in the testes, ovaries, and spermathecae (Beckmann et al. 2013; Beckmann and Fallon 2013).

Unfortunately, the iTRAQ experiment produced very little reliable data reflected up or down regulated host proteins. However, it did produce a significant list of Wolbachia expressed proteins and a relatively quantitative means of assessing their abundance within the ovary in comparison to one another. The iTRAQ produced *Wolbachia* proteome correlates and validates our previous data. Table 1 shows 78 *Wolbachia* expressed proteins. Proteins were reported only if three peptides were detected from that particular protein with (95%) probability each. The proteins were then sorted according to their relative abundance as determined by the number of spectra from the single most abundant peptide; we used this as an estimate of protein abundance. Because proteins can often produce varying amounts of detectable tryptic peptides depending upon protein size and lysine/arginine content, to control for these variable aspects we reasoned that most proteins would be able to produce at least one detectable tryptic peptide and counting the quantity of the single most abundant peptide would then be a better reflection of protein concentration rather than total peptides derived from that protein. This and previous data justifies our quantification because it reports that the top two abundant proteins are the *Wolbachia* surface protein (WSP; gi|190571332) and another putative membrane protein (gi|190570988) (Table 1); our previous study showed that these same two proteins produced the only clearly visible SDS-PAGE bands in gels

comparing both infected/uninfected testes and ovaries (Beckmann et al. 2013; Chapter 3). Here we also repeat detection of the HU beta protein (gi|190571020) within the ovaries, its abundance is ranked as a 2 compared with that of the WSP and putative membrane protein which are 42 and 29 respectively. As expected, other highly expressed proteins are from groups of proteins containing surface protein components, chaperones, and ATPase synthase alpha subunit.

Interestingly, this ovarian *Wolbachia* proteome also corroborates our other proteomic work performed within the spermathecae, where we detected the presence of WPIP0282 (gi|190570728) in mosquito sperm enriched protein extracts. These data show that the relative abundance of this protein falls very high on the list at 8 with few other proteins being more abundant. In Chapter 4 we proposed that this protein is a candidate CI inducer and also suggested that any such protein would likely be expressed in an extremely abundant fashion because CI correlates with *Wolbachia* density in the testes. These data support that initial hypothesis and show that WPIP0282 is a highly abundant protein. We also showed that wPip_0282 is expressed as an operon with *w*Pip_0283 and has itself been duplicated additionally within the *w*Pip genome to *w*Pip_0294 and *w*Pip_0295. All four of these proteins appeared within our proteome. In addition to WPIP0282, WPIP0283 and WPIP0294 were detected with relative abundances of 2 and WPIP0295 was detected with only 1 total (95%) peptide. These data suggest that of the two putative CI-inducing *Wolbachia* operons, wPip_0282/0283 is more highly expressed than wPip_0294/0295, because its relative abundance (determined by the spectral counting described above) was far higher than those of the alternative operon paralog.

	Abundance	Name	Species	Accession #	Total Peptides (95%)
*	42	surface antigen Wsp	wPip	gi|190571332	128
*	29	Putative membrane protein	wPip	gi|190570988	68
	15	chaperone protein dnak (hsp70)	wPip	gi|190570602	32
	12	chaperonin, 60 kDa	wPip	gi|190570503	94
	10	ATP synthase F1, alpha subunit	wPip	gi|190571573	75
	9	chaperonin, 10 kDa	wPip	gi|190570502	24
*	8	WPIP0282	wPip	gi|190570728	68
	8	peptidoglycan-associated lipoprotein, putative	wPip	gi|190571199	20
	7	Putative outer membrane protein	wPip	gi|190571111	59
	7	translation elongation factor tu	wPip	gi|190571544	25
	6	ankyrin repeat domain protein	wPip	gi|190570819	18
	6	putative phage related protein	wPip	gi|190570849	9
	6	putative phage related protein	wPip	gi|190571703	12
	5	Hypothetical protein WP0984	wPip	gi|190571376	37
	5	Hypothetical protein WP0890	wPip	gi|190571287	23
	5	ribosomal protein L16	wPip	gi|190571553	6
	4	translation elongation factor G	wPip	gi|190570976	14
	3	polyribonucleotide nucleotidyltransferase	wPip	gi|190571231	16
	3	ribosomal protein L7/L12	wPip	gi|190570969	15
	3	putative phage related protein	wPip	gi|190571688	5
	3	hypothetical protein WP0171	wPip	gi|190570629	4
	3	thioredoxin	wPip	gi|190571104	5
	3	recA protein	wPip	gi|190571327	4
	2	DNA-directed RNA polymerase, beta/beta' subu	wPip	gi|190570968	25
	2	surface antigen	wPip	gi|190571424	8
*	2	DNA-binding protein, HU family	wPip	gi|190571020	10
	2	enhancing lycopene biosynthesis protein 2, put	wPip	gi|190571210	8
	2	DNA-directed RNA polymerase, alpha subunit	wPip	gi|190571569	7
*	2	WPIP0283	wPip	gi|190570729	9
	2	ribosomal protein S1	wPip	gi|190571429	8
*	2	WPIP0294	wPip	gi|190570737	6
	2	translation elongation factor Ts	wPip	gi|190571620	5
	2	transposase	wPip	gi|190571636	6
	2	antioxidant, AhpC/Tsa family	wPip	gi|190570611	5
	2	Putative dnaj domain membrane protein	wPip	gi|190570961	4
	2	ribosomal protein S5	wPip	gi|190571563	4
	2	conserved hypothetical protein	wPip	gi|190570734	5
	2	ribosomal protein L1	wPip	gi|190570971	4
	2	transcription elongation factor GreA	wPip	gi|190571574	4
	2	cold-shock domain family protein	wPip	gi|190571462	5
	2	50S ribosomal protein L11	wBm	gi|58584908	3
	2	Hypothetical protein WP0985	wPip	gi|190571377	3
	2	translation initiation factor IF-2	wPip	gi|190571749	2
	2	30s ribosomal protein s6	wPip	gi|190571063	2
	2	putative phage related protein	wPip	gi|190571691	7
	2	putative phage related protein	wPip	gi|190571690	7
	2	Phage related DNA methylase	wPip	gi|190571683	4
	2	ompA-like protein	wPip	gi|190571144	5
	2	two component transcriptional regulator	wPip	gi|190570997	3
	2	ribosomal protein S3	wPip	gi|190571552	4
	2	superoxide dismutase, Fe	wPip	gi|190571001	3
	1	putative phage related protein	wPip	gi|190571689	6
	1	protease DO	wPip	gi|190571439	5
*	1	WPIP1117	wPip	gi|190571499	6
	1	membrane GTPase involved in stress response	wBm	gi|58584322	6
	1	ribosomal protein S4	wPip	gi|190570680	4
	1	N utilization substance protein A	wPip	gi|190571750	4
	1	transcription termination factor Rho	wPip	gi|190570947	5
	1	heat shock protein HtpG	wPip	gi|190571174	3
	1	ribosomal protein L2	wPip	gi|190571549	3
	1	trigger factor, putative	wPip	gi|190570981	3
	1	pyruvate dehydrogenase complex, E3 compone	wPip	gi|190570560	4
	1	succinyl-CoA synthase, beta subunit	wPip	gi|190571356	3
	1	peptidase, M16 family	wPip	gi|190570922	3
	1	ribosomal 5S rRNA E-loop binding protein Ctc/L	wPip	gi|190571325	3
	1	HIT family protein	wPip	gi|190571250	2
	1	Hypothetical protein WP0828	wPip	gi|190571228	1
	1	cytosol aminopeptidase	wPip	gi|190571477	1
	1	phosphoribosylamine--glycine ligase	wPip	gi|190570964	4
	1	ribosomal protein S2	wPip	gi|190571619	3
	1	hflC protein	wPip	gi|190571440	4
	1	hypothetical protein WP0593	wPip	gi|190571002	3
	1	iron compound ABC transporter, periplasmic irc	wPip	gi|190571080	3
	1	ribosomal protein L28	wPip	gi|190570684	3
	1	ribosomal protein L3	wPip	gi|190571546	3
	1	hypothetical protein WP0065	wPip	gi|190570536	3
	1	bacterioferritin comigratory protein	wPip	gi|190571297	3
*	1	WPIP0295	wP{ip	gi|190570738	1

Table 1. Total Wolbachia expressed proteins from infected ovaries as detected in the iTRAQ experiment. Proteins are listed according to their relative abundance, calculated as described in the text. Proteins highlighted in red with an asterisks were detected and discussed in the previous mass spectrometry experiments described here in Chapters 3 and 4. The table both verifies detection of previously detected proteins within the ovaries and demonstrates that both operon copies of wPip_0282/0283 and wPip_0294/295 are also expressed within the ovaries. Specifically, I note that WPIP0282 is an extremely abundant protein expressed on comparable levels to some surface proteins.

We also previously proposed that the operon likely behaves as a toxin-antidote system. It is often the case, that in these systems the antidote is less stable than the toxin and toxic effects occur when the antidote degrades and transcription of the operon ceases. In each operon expression of the first gene is always more abundant than the second protein. This is unlikely to be the case were the proteins of equal stability because both WPIP0283 and WPIP0295 are much larger than their operon partners WPIP0282 and WPIP0294 and consequently should produce more tryptic peptides. I argue that because the second genes of each operon are present within the proteome at lesser relative abundances than the first genes the putative antidote proteins are possibly less stable than the putative toxins. This study is the first proteomic verification of protein production from both operons encoding all four genes in *Wolbachia* infected ovaries.

CHAPTER 5:

CLONING and CHARACTERIZATION of the *WOLBACHIA* SPERMATHECA ASSOCIATED PROTEIN WPIP0282 in a HIGHLY REPRESSIBLE EXPRESSION SYSTEM for POTENTIALLY TOXIC PROTEINS

Introduction:

The previous chapters focused on developing and utilizing methods to search for *Wolbachia* expressed proteins contained within infected and uninfected *Culex pipiens* mosquitoes, specifically within their reproductive tissues. My attention soon became focused upon the WPIP0282/0283 operon, whose protein products were shown to accumulate within *Wolbachia* modified sperm and ovaries. Because these proteins are the only proteins that have been shown to be in the right place at the right time, they were hypothesized to be prime candidates potentially responsible for induction and rescue of CI. Further bioinformatics analysis justified focused attention on this operon. The operon seemed to be organized in a way similar to toxin-antidote operons, was completely absent in strains of non CI-inducing *Wolbachia*, and the second gene in the operon, WPIP0283, contained a catalytically conserved C-terminal SUMO protease domain which was putatively of eukaryotic origin. In order to further continue researching this unique operon I undertook the task of cloning these two *Wolbachia* genes, first starting with *w*Pip_0282.

The task of cloning the *w*Pip_0282 gene proved extremely difficult and spanned the course of two years. My original hypothesis, that WPIP0282 acts as a toxin, seemed

to be verified in the course of these experiments. Initial attempts to clone the protein in the pTrcHis vector only resulted in the generation of reverse oriented inserts. Further experimentation led us to believe that *E. coli* preferred not to maintain plasmids expressing the *w*Pip_0282 sequence so we switched our vector system to one designed for the cloning and expression of toxic proteins, pBAD (Invitrogen). Cloning attempts within pBAD produced mutant clones, which appeared to be mutated at conserved protein residues, thereby limiting our ability to test any functionality of the protein. Further attempts to correct the mutated *w*Pip_0282 failed.

Moderately toxic genes are defined as genes that negatively interfere with the normal physiology of *E. coli* during the induction phase only. In contrast, highly toxic genes interfere with the physiology of *E.coli* during the growth phase (Saida, 2007). Oftentimes, leaky expression of repressed genes can be sufficient to prevent the survival of transformed bacterial cells. In such cases, there are various strategies that can be used to super-repress these particular genes. Some of these strategies include phage mediated delivery of high specificity RNA polymerases, tightly regulated promoters, as well as antisense and competitive promoters (Saida et al., 2006). I chose to use an antisense competitive promoter to induce constant transcription of antisense transcript because it had previously been described as a successful strategy to clone a bovine toxic DNase I protein (Worrall and Connolly, 1990). Only by re-engineering the pBAD vector to a super-repressed state was I able to clone and express wildtype *w*Pip_0282.

In addition I also cloned and expressed the catalytic SUMO protease domain of WPIP0283 so that in the future I would be able to test its enzymatic cleavage of SUMO substrates. In contrast to the cloning of *w*Pip_0282, cloning of the SUMO protease domain was extremely easy with the normal pBAD vector. Functional testing of these proteins is currently ongoing.

Materials and Methods:

DNA extraction. *Wolbachia* genomic DNA was extracted as described by Beckmann and Fallon, 2012. Decapitated mosquitoes were individually homogenized in 200 µl of 120 mM Tris-HCl, pH 9, containing 0.5% SDS, 80 mM NaCl, 160 mM sucrose and 60 mM EDTA. After 30 min at 65°C, potassium acetate (28 µl of 8 M) was added, mixed by vortexing, and the sample was incubated on ice for 30 min. Samples were centrifuged for 10 minutes in a microcentrifuge at 15,000 rcf, and the resulting supernatant (180 µl) was placed into a new 1.5 ml microcentrifuge tube and 360 µl of 100% ethanol was added. The samples were then briefly vortexed and held overnight at -20°C. Nucleic acids were pelleted by centrifugation at 15,000 rcf for 10 min and the pellets were dried under vacuum, before resuspension in 10 mM Tris-HCl, pH 7.5, containing 0.4 M NaCl and 10 µg boiled RNAseA (400 µl) at 37°C for 1 h. Samples were extracted with an equal volume of phenol, and the aqueous phase (380 µl) was transferred to a new microcentrifuge tube. The phenol phase was re-extracted with 400 µl of 10 mM Tris-HCl, pH 7.5, containing 0.4 M NaCl, and the combined aqueous phases were precipitated with ethanol overnight at -20°C. DNA was recovered by

centrifugation, washed in 70% ethanol, dried and dissolved in 100 µl double-distilled water.

Polymerase Chain Reactions and Primers. All polymerase chain reactions were performed with an initial denaturation step for 5 minutes at 95°C, followed by 35 cycles of three consecutive steps of: denaturation at 95°C, annealing at the corresponding temperature (Table 1), and extension at 72°C. The duration of denaturation and annealing steps were set to 1 minute, while extension times were varied based on the size of the PCR amplicon (1 minute/kb). Primers 0282BstBIF/R were used for non-directional cloning of wPip_0282 into the pTrcHisC vector (Invitrogen). Primer pairs 0282XhoIF/0282BstBIR and 0282BstBIF/ 0282XhoIR were used for directional cloning of wPip_0282 in the forward and reverse orientations respectively into pTrcHisB. The primer pair 0282XhoIF/0282EcoRIR was used for forward orientation cloning of wPip_0282 into the pBADB vector (Invitrogen) and the re-engineered Anti-pBAD vector. Primers LacBstBIF/R were used to clone the regulatory region from the pTrcHis vector into the pBADB vector. Primer pairs 0282XhoIF/0283EcoRIR were used in attempts to clone the entire wPip_0282-0283 operon into pBADB. The Primer pair 0283IntXhoIF/0283IntBstBIR was used to clone the C-terminal SUMO protease region of wPip_0283 into pBADA vector.

Transformation, Selection, Screening, Strains, and Sequencing. Transformations of *E. coli* strains were carried out according to the pTrcHis manual by Invitrogen. Briefly bacteria were grown to 0.5 OD_{600} and transformed using 50 mM $CaCl_2$ treatment and heatshock. Transformed bacteria were allowed to grow for ~~one hour~~

1 h in SOC medium and were then plated onto 50 μg/ml Ampicillin LB plates for selection. Colony forming units were then grown in LB Amp media as in the case of Top10 *E. coli* or RM Amp media as in the case of LMG194 (expression strain) *E. coli*. In the case of transformations and growth of bacteria with the Anti-pBAD vector, IPTG was always present in the medium and plates at a concentration of 1 mM. Plasmids were extracted by the bacterial alkaline lysis mini prep procedure described in the Invitrogen pTrcHis manual. Plasmids were screened for inserts by assessing gel mobility shift on an 0.8% agarose gel. Highly purified plasmid for sequencing was extracted using the High Pure plasmid Isolation Kit from Roche (Indianapolis, IN).

Recombinant Protein Expression, Extraction, and Tandem Mass Spectrometry. Recombinant protein expression was induced in the case of pBAD by induction with a final concentration of 0.2% arabinose when the bacteria reached 0.5 OD_{600}. In the case of Anti-pBAD clones, expression of recombinant proteins were induced by centrifugation of the transformed bacteria, removal of medium containing IPTG, and replacement with medium containing 0.2% arabinose. Recombinant his-tagged protein was then extracted by pulldowns with the Dynabead His-Tag Isolation and Pulldown kit from Life Technologies using the manufacturer's specifications. Protein samples were reconstituted in SDS sample buffer and boiled prior to electrophoresis, which was conducted on 8-18% gradient polyacrylamide gels. Protein gels were submitted to the University of Minnesota's Center for Mass Spectrometry and Proteomics for in-gel chymotrypsin digestion and tandem mass spectrometry.

Primer	Sequence	Annealing	Target Gene
0282**BstBI** F	TGGGAATTCGAACTTAATAGGGGGGATAATGCC	59	wPip_0282
0282**BstBI** R	TGGGAATTCGAACCATTACTCATACAAAATACCCC	59	wPip_0282
0282**XhoI** F	TGGGAACTCGAGCTTAATAGGGGGGATAATGCC	59	wPip_0282
0282**XhoI** R	TGGGAACTCGAGCCATTACTCATACAAAATACCCC	59	wPip_0282
0282**EcoRI** R	AACCAAGAATTCCCATTACTCATACAAAATACCCC	59	wPip_0282
0283**EcoRI** R	AAGGAAGAATTCGACCGATTACAACCAATGCG	59	wPip_0283
Lac**BstBI** F	CGTTAATTCGAACAGACCGTTTCCCGCGTG	60	LacO; Ptr; LacI;
Lac**BstBI** R	CGTTAATTCGAAGGTCTAGTGCCCACACAG	60	LacO; Ptr; LacI;
0283Int**XhoI** F	TGGGAACTCGAGGACATGTTTTTCGCTGCACG	58	wPip_0283
0283Int**BstBI** R	TTTTTTTTCGAAGACCGATTACAACCAATGCG	58	wPip_0283

Table 1. Primers used for cloning purposes within this study. Annealing temperatures are given in degrees Celsius. Red residues are restriction enzyme sites.

Results:

WPIP0282 is recalcitrant to cloning in E. coli.

pTrcHisC vector: Initially, simple non-directional cloning of *w*Pip_0282 into the pTrcHisC vector was attempted. PCR products were produced using forward and reverse primers, each with a terminal BstBI restriction enzyme site. We sequenced the PCR product to confirm its sequence and that the BstBI sites were being added onto the products properly. The sequence of *w*Pip_0282 (Buckeye) was published in Beckmann and Fallon, 2013 (see Chapter 4). Upon transformation into competent Top10 *E.coli,* we produced 40 ampicillin resistant clones of which only one clone had the *w*Pip_0282 gene insert in the reverse orientation. We then attempted to simply flip the reverse oriented *w*Pip_0282 gene by re-digestion, re-ligation, and re-transformation. After transformation we picked 72 colonies for growth and screening. Of those 72 only 21 clones survived to be cultured in liquid medium and only 1 of those 21 had the *w*Pip_0282 insert; again in the reversed orientation.

Because we suspected that *w*Pip_0282 might be the CI toxin we assumed that it might also be toxic to *E.coli*. We then re-designed the primers with different restriction enzyme sites on each end of the *w*Pip_0282 gene in order to force directional cloning into the pTrcHisB vector in both the forward orientation (XhoIF/BstbIR primers) and as a positive control, the reverse orientation (BstbIF/XhoIR primers). We hypothesized that if the protein was toxic we would be able to clone it again in the reverse orientation, but not in the forward orientation; this experiment would also rule out procedural mistakes and skill as a limiting factor. The PCR products were again sequenced to verify restriction

site addition. Competent Top10 *E. coli* cells were then transformed separately with the forward and reverse ligated *w*Pip_0282/vector constructs. One hundred AMP resistant colonies were obtained from the forward orientation transformants whereas 88 AMP resistant colonies were obtained with the reverse orientation construct. Colonies (24 each of both forward and reverse construct transformations) were picked, and all 48 clones survived until plasmid extraction. Of the 24 forward orientation clones none turned out to have correct inserts and in fact, some looked as if they had been rearranged as determined by restriction digestion (data not shown). In contrast, of the 24 positive control, reverse orientation clones, one had the *w*Pip_0282 gene insert. Because we were again able to successfully clone the gene backwards but not forwards we concluded that the protein was toxic and could not be cloned in the forward orientation with our current pTrcHis vector. The pTrcHis vector system is driven by a strong tryptophan promoter and controlled by the lac repressor and operator. The lac system operates in a way that is constitutively on, but always repressed by the lac repressor, without IPTG. We assumed that this system was too strong and leaky for expression of putatively toxic proteins.

Switching to the pBAD Vector System. We then sought to clone *w*Pip_0282 in a vector specifically designed to clone toxic proteins and hold leaky transcription of inserted genes at a bare minimum. I used the pBAD vector system because it has been shown to be valuable for toxic proteins and is tightly regulated by the arabinose system (Lee et al., 1987). We produced new PCR *w*Pip_0282 gene products with the (XhoIF/EcoRI primers), digested, and ligated them into the pBADB vector. In these initial experiments we observed even fewer colonies than with the pTrcHisB vector (data

not shown). Three of 36 AMP resistant colonies which were successfully cultured and screened were determined to have the insert by plasmid separation on an agarose gel. The three clones, #9, #23, and #42, were then sequenced. Both clones, #9 and #23 had point mutations causing changes in Amino Acid sequence (Fig. 1). Intriguingly, these mutations, #9:L425S and #23:R401G uniquely appeared to fall on completely or mostly conserved residues, suggesting a change in functionality of the protein. The third clone #42 turned out to be a very similar, but divergent homolog of wPip_0282 occurring in *w*Pip(JHB), (JHB1: ZP_03335681.1; Beckmann and Fallon, 2013), but was mutated with a premature stop codon truncating the C-terminus. This result indicates that our *w*Pip(Buckeye) strain has at least 3 homologous copies of *w*Pip_0282, with multiple copies being amplified in our PCR reactions, further compounding our efforts to retrieve a perfect wild type copy of specifically the *w*Pip_0282 Buckeye paralog. Subsequent transformation experiments provided two more mutants, again with mutations in highly conserved residues; clone #15 with a K245R mutation and clone #4 with a G407D mutation. Because the sequences of these four clones each contained at least one mutation diverging from the wildtype sequence in a conserved residue, and we were able to successfully clone the perfect sequence in reverse orientation multiple times, we hypothesized that the mutations arose in *E. coli* itself, and were not a result of our Taq Polymerase. However, it is difficult to eliminate the possibility that a small portion of the amplified product contained these conserved residue mutations which then offered an advantage to bacteria expressing mutant protein as opposed to wildtype protein.

We performed further experimentation with the cloning and transformation procedures in attempts to yield a perfect clone. We tried direct cloning into the special LMG 194 *E. coli* expression strain provided with pBAD, included glucose in our medium (which represses pBAD further), and varied the temperatures from 23-37 °C, without success. To test our hypothesis of the toxin-antidote operon we also tried cloning the entire ~ 4500 bp *w*Pip_0282/0283 operon as a whole. This attempt was also unsuccessful. I also tried unsuccesssfully to use a GeneEdit site directed mutagenesis kit to fix clone #15's K245R mutation. Finally, I tried utilizing both clone #15, which has a perfect C-terminus and Clone #23 which has a perfect N-terminus to construct a perfect copy by digestion with BglII (a site that falls between the two mutations), gel purification, and re-ligation. As a positive control I used this strategy to easily construct the double mutant from the corresponding BglII fragments. In contrast, transformation with the wildtype in vitro recombinant produced only two transformants, of which neither had the insert. Table 2 shows a summary of all these experiments and their results.

Re-engineering the pBAD Vector and Cloning of Wildtype WPIP0282. As a final attempt to clone the wildtype sequence of *w*Pip_0282 I sought to re-engineer the pBAD vector in a way that would make it super-repressed. I reasoned that the inability to clone *w*Pip_0282 within an expression vector was a direct result of both vectors (pTrcHis and pBAD) being leaky and allowing minimal unwanted expression of WPIP0282. To circumvent this problem I used the lacBstBIF/R primers to amplify 3 regulator elements from pTrcHis and insert them into the 3' end of the multi-locus cloning site of pBAD in an antisense direction using the BstBI restriction enzyme (Figure 2). The regulatory

```
                                        #15
                                         R
wPip_0282       222   APILDFCVNKIVDKDTL-LQKLSQKDKGVYSLFAELIESCFFDTVHDLVQCWCYKEVSAG
wMel_0631       218   APILDFCVRNIGDKDTL-LQKLLQKDKGVYFLLAELIESCFFDTVHDLVQCWCYKGVSAG
wRi_005370      218   APILDFCVRNIGDKDTL-LQKLLQKDKGVYSLLAELIESCFFDTVHDLVQCWCYKGVSAG
wPip_0294       226   VDTIEFCLSKMDDEQKKKLLDRDYKENTYYAVLNVLVGQYYFDSFMELSRLC--------
                      .  ::**: :: *::.  * .   *::  * ::  *: . :**:. :* :


wPip_0282       281   GDHSEKIFSQRDYELFLSSLSDVMLKNPESNVQARSLIMEFWECGSLYQYRKAAVNTSNY
wMel_0631       277   GDCSDKIFSQQDYELFLYSLSNVMLKNPELSVQARSLIMEIWKCERFAEYRETSVNTSNY
wRi_005370      277   GDCSDKIFSQQDYELFLYSLSNVMLKNPELSVQARSLIMEIWKCERFAEYRETSVNTSNY
wPip_0294       278   -----SQIECERYTTFLSSLSDQVLKNPDLSEETKKCMMNVWERIIKLKTQDRGE-----
                           . :. . *  ** ***: :****: . :::. :*:.*:     : :. .


                                                      #9
                                                       L
wPip_0282       341   TVPTSGVFAELIVNWRREDIYKTDEEKEIEKKEILDMMSF-AKDCFPEKFELFKKLIIRD
wMel_0631       337   TVPIKSVLGGLIINWKREDVCKPD--REIEKEEILDMISF-AKGCFPEKFDLFKEVMIEN
wRi_005370      337   TVPIKSVLGGLIINWKREDVCKPD--RKIEKEEILDMISF-AKGCFPEKFDLFKEVMIEN
wPip_0294       328   -QSISSIFVDYSVTYTIANLIVDPSRQGVSKEEILGKILKHVKEMSGEEMIKVKDSVLSK
                          ..::    :.:   ::      : :.*:***  :   .*    *::  .*. :: .


                      #23  #4              #9
                       G    D               S
wPip_0282       400   LRLCGREGKRVNVDYGLFAEELFSELEKTILPPGPVGDGPCSNLRSRSKAHGSKKTTLPV
wMel_0631       394   LRICGREGKRKGVDYGKFAEELFLQLEKVTLP--SVGDGPWNNLRSQSKVSLP---LDGS
wRi_005370      394   LRICGREGKRKGVDYGKFAEELFLQLEKVTLP--SVGDGPWNNLRSQSKVSLP---LDGS
wPip_0294       387   IQLF-HGGKKLQ-----LGEQVFSKLAQEASKESI--------LR------------EAG
                      :::  : **:       :.*::* :* :               **
```

Figure 1. Amino acid alignment of residues 222-458 from WPIP0282 (wPip Buckeye) and its corresponding homologs in *w*Mel, *w*Ri; *w*Pip_0294 is a divergent paralog in *w*Pip *quinquefasciatus*. Asterisks indicate fully conserved residues, two dots indicate strongly conserved residues, single dots indicate moderately conserved residues. Black boxes and residues indicate the clone number and mutation/s contained in that individual clone. The Alignment was made in ClustalW Omega.

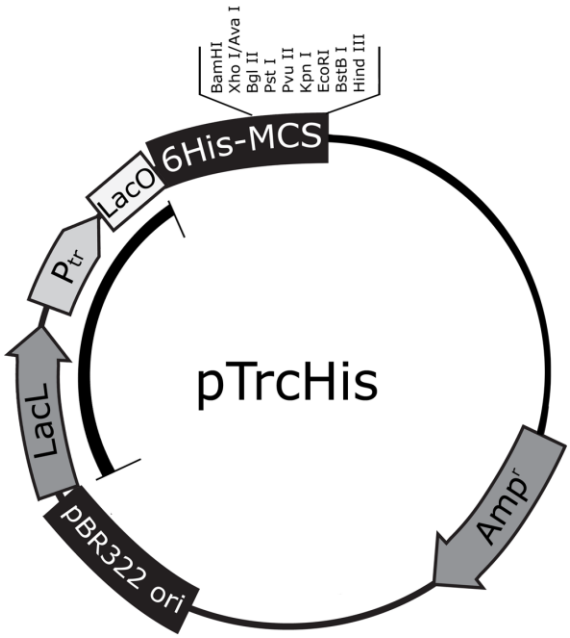
Exp	Vector	Technique	# AmpR	# Survivors	# with Insert	Orientation	Clone #
1	pTrcHis	non-directional cloning	40	37	1	Reverse	
2	pTrcHis	Flipping reverse clone	72	21	1	Reverse	
3	pTrcHis	Forward Cloning	24	24	0		
4	pTrcHis	Reverse Cloning	24	24	1	Reverse	
5	pBAD	Forward Cloning	36	36	3	Forward	4,9,15,23,42
6	pBAD	Cloning of wPip_0282/0283 operon	20	17	0		
7	pBAD	Site-Directed Mutagenesis of #15	24	24	2		
8	pBAD	Ligation of #15 and #23					
9	Antisense-pBAD	Forward Cloning	24	9	9	Forward	L10 (wildtype)

Table 2. Summary of experiments undertaken to clone the wildtype WPIP0282

elements added were the lac operator, LacO, which is a binding site for the lac repressor, LacI, whose gene was also added and the tryptophan promoter. These elements were added in an antisense directionality to induce expression of antisense transcript in the presence of IPTG. The vector would then be super-repressed in the sense that it would be repressed not only by the arabinose regulated AraC repressor, but also by the continuous production of antisense transcript. When the operon needed to be turned on, then one simply removed IPTG and added arabinose.

Interestingly, the re-engineered vector was immediately successful and I was able to clone multiple wildtype *w*Pip_0282s in initial trials (Clone #L10). Even more surprising was the fact that once inside the new vector, *E. coli* did not seem to mind expressing the putatively toxic protein (Figure 3). Growth studies measuring OD_{600} in LB medium confirmed that while expression of WPIP0282 slows growth minimally, there is no substantial effect or death as one would expect from expression of a toxic protein (data not shown). However, in the presence of arabinose the protein expresses as a double band (Figure 3, compare lanes 2 and 3.) The His-tagged WPIP0282 has a molecular mass of approximately 60.6 kDa, consistent with the top band of the doublet. Interestingly, when we first detected the protein in spermathecae, it also exhibited a distinct double banding pattern (Beckmann and Fallon, 2013). It is unclear whether or not cleavage within E. coli, affects the functionality and or putative toxicity of the protein.
Both mutant clones #23 and #15 also expressed as protein doublets. Prior to successfully cloning a wildtype copy of the protein I performed extensive proteomic analysis of these doublet bands from clone #15 K245R mutant in hopes to determine the nature of the

cleavage. Logically the protein could not be cleaved on the N-terminal end as this would remove the His-tag and the cleaved product would not pull down with the cobalt beads. Likewise, cleavage could not be in the middle of the protein or the cleaved products would migrate significantly lower on the gel. However, a C-terminal cleavage site seemed to be most consistent with the approximately 7 kDa shift in mobility of the lower band observed in Fig. 3. I excised both the top and bottom bands from gels of expressed product consistent with the gel shown in figure 3 and subjected them to digestion by chymotrypsin and peptide analysis in the Peaks 6 program. I hypothesized that C-terminal peptides from the cleaved bottom band would be less abundant than those from the top bands. Both bands were clearly composed of WPIP0282 as mass spectrometry analysis of the top band detected 303 peptides covering 94% of the protein and analysis of the bottom band detected 265 peptides covering 96% of the protein. Peptides derived from C-terminus of WPIP0282 in the bottom band showed a 3-fold drop in detected spectra when compared to those quantities from the top band (data not shown). This data supported my hypothesis of C-terminal cleavage however a repeat of the experiment with a different enzyme, trypsin, did not show the same pattern. The nature of the cleavage is currently unknown.


BamHI
Xho I/Ava I
Bgl II
Pst I
Pvu II
Kpn I
EcoRI
BstB I
Hind III
6His-MCS
LacO
Ptr
LacI
pBR322 ori
Ampr
pTrcHis

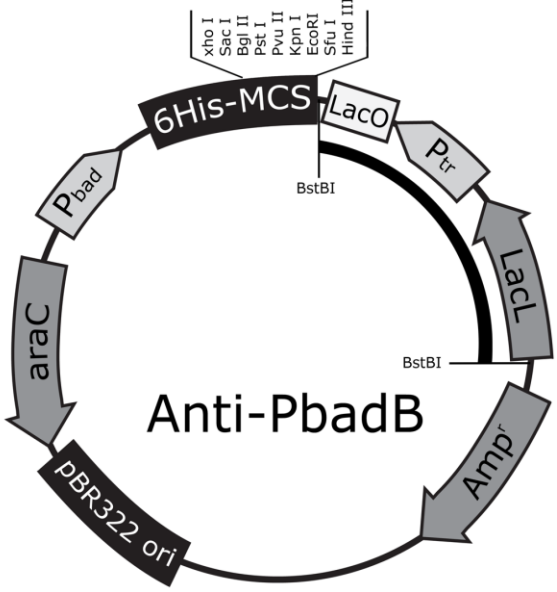

Xho I
Sac I
Bgl II
Pst I
Pvu II
Kpn I
EcoRI
Sfu I
Hind III
6His-MCS
LacO
Ptr
BstBI
Pbad
araC
LacI
BstBI
pBR322 ori
Ampr
Anti-PbadB

Figure 2. pTrcHis and Anti-Pbad Vectors. LacL is the Lac repressor ORF, Ptr contains the -35 region of the trp promoter together with the -10 rgion of the lac promoter, LacO is the lac operator site (binding site of LacL repressor), araC is the arabinose binding *ara* regulator. Pbad is the arabinose operon *ara*BAD promoter. A PCR amplicon was created from amplifying the regulatory region from pTrcHis and was ligated 3' of the multiple cloning site, MCS, of the pBADB vector to create the *Anti-PbadB* vector.

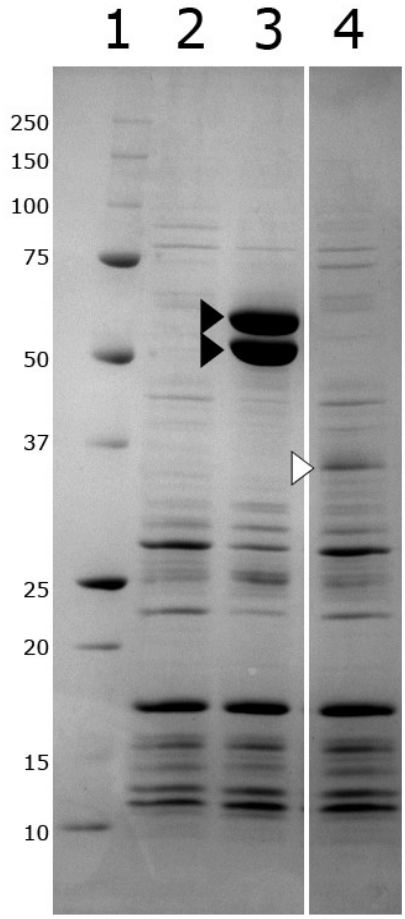

1
2
3
4
250
150
100
75
50
37
25
20
15
10

Figure 3. SDS-PAGE gel analysis of his-tagged recombinant proteins His-WPIP0282 from the perfect clone L10 and His-WPIP0283 (891-1174) from clone K2. Lane one is a kDa marker, lane two is un-induced clone L10, lane three is induced (+ arabinose) L10, lane four is induced (+ arabinose) K2. The two black triangles show the double banding pattern of the 60.6 kDa His-WPIP0282. The white triangle shows the 36.1 kDa His-WPIP0283 (891-1174).

Before beginning intensive functional analysis of the wildtype WPIP0282 I thought it necessary to also clone *w*Pip_0283 such that I could perform experiments with both proteins together and test their interactions. Because WPIP0283 is an extremely large protein of ~1500 amino acid residues, I thought it best to first clone and express a truncated construct only containing the C-terminal fragment with the SUMO protease domain. Other researchers studying similar Ulp1 SUMO proteases have noted that the catalytic C-terminus is sufficient for enzymatic cleavage and that the N-terminal and middle portions of the proteins mostly regulate localization (Li and Hochstrasser, 2003). The *w*Pip, endosymbiont of *Culex quiquefasciatus Pel,* strain also encodes a homologous truncated copy of this SUMO protease domain in the gene *w*Pip_1291 (Beckmann and Fallon, 2013). Thus, in essence I was cloning both *w*Pip_1291 and the C-terminal region of *w*Pip_0283. In contrast to the nearly intractable *w*Pip_0282, cloning of the C-terminal region of *w*Pip_0283 produced wild type clones in initial experiments. Expression of this recombinant protein is shown in Lane 4 of Fig. 3. I note that the C-terminus of second protein WPIP0283 is expressed as a single band, but levels of expression are less robust than the doublet WPIP0282 bands. These observations suggest that cleavage of the putative toxin, WPIP0282, is due to something unique about its sequence or function, rather than due to random proteases within the *E. coli* expression strain LMG194.

Discussion:

My attempt to clone and characterize the *Wolbachia* spermatheca associated protein, WPIP0282, was an extensive process involving sequential transitioning to three different cloning vectors, each one being more tightly regulated than the last. Only by cloning the gene within an extremely controllable and tightly repressed construct, Anti-Pbad, was I able to obtain an expressible wildtype copy. However, successful cloning of the wildtype sequence in the Anti-Pbad construct only created more questions. The protein, which had previously been assumed to be toxic, did not seem to display any harmful effects on *E. coli* aside from mildly retarded growth. One hypothesis explaining the inconsistencies of these results is that the double banding pattern observed in the expression of the recombinant His-WPIP0282 is the result of a digestion product of the protein which may have reduced functionality. We note that even the top band, which is putatively un-cleaved, migrates slightly lower than the expected 60.6 kDa. This opens the possibility that the protein might even be cleaved twice or more. However the mysterious double banding pattern remains unsolved at this point.

My original proposal of WPIP0282 being a CI toxin inherently implied an antidote functionality to the protein WPIP0283. Another alternative hypothesis explaining the seemingly non-toxic nature of WPIP0282 is that *E. coli* might itself have a derivative antidote or have selected for a complementing mutation in some other gene. However, BLAST analysis of the *E. coli* genome shows no proteins homologous to WPIP0283 or any proteins with Ulp1-like SUMO protease domains. Also, cloning of the entire operon did not result in the generation of a wild type clone as one would expect if I

had also cloned the antidote, however, these results are still under investigation and subsequent attempts to clone the operon did result in additional clones that have not been sequenced, so it would be premature to suggest that these results rule out the toxin-antidote hypothesis of the *w*Pip_0282/0283 operon. On the other hand the inability to directly measure any toxicity of WPIP0282 within *E. coli* is a significant criticism of these data and will need to be further understood before publication of these results. I believe that further testing of the WPIP0282 protein as well as the catalytic SUMO protease domain of WPIP0283 will shed light on these seemingly contradictory results. In preparation for continued investigation during my upcoming postdoctoral research I have already begun the process of testing WPIP0283 (residues 891-1174) for its ability to cleave *Culex* SUMO and also its ability to interact with WPIP0282. If WPIP0283 is demonstrated to have the ability to cleave insect SUMO this will be an important discovery regardless of the operons function with respect to CI.

CHAPTER 6:

FINAL HYPOTHESES and IN-SILICO TESTING OF THE WPIP0282/0283 TOXIN-ANTIDOTE THEORY OF CYTOPLASMIC INCOMPATIBILITY

Introduction:

The previous chapter detailed extensive cloning experiments which seemed to support the hypothesis that WPIP0282 acted as a toxin within *E.coli* and therefore was possibly the CI toxin in *Culex*. However, once the final wildtype copy of WPIP0282 was obtained, toxic effects of the protein could not be measured within the bacteria; leaving the *w*Pip_0282/0283 toxin-antidote hypothesis of CI unresolved. However, recent *Wolbachia* genome publications provided a unique opportunity to test my CI theory. Here I discuss in-silico genomic comparisons which were used to support the toxin-antidote hypothesis of the *w*Pip_0282/0283 operon with the logic being thus: because the two *Wolbachia* strains *w*Pip(Pel) and *w*Pip(Mol) are incompatible, if the operon induces CI it is necessary that the those genes be in some manner different between the two genomes of extremely closely related *Wolbachia*.

The strains *w*Pip(Pel) and *w*Pip(Mol) (henceforth: *w*Pel and *w*Mol) infect *Culex pipiens quinquefasciatus Pel* and *Culex pipiens molestus* respectively. The two mosquitoes are considered to be the same species and are morphologically indistinguishable, however they are reproductively incompatible (Byrne and Nochols,

1999). This is reportedly due to bi-directional incompatibility induced by divergences in the genomes of their corresponding *Wolbachia* strains, *w*Pel and *w*Mol (Pinto et al., 2013). In essence, each *Wolbachia* strain cannot rescue the others' reproductive sperm toxin. Recently, the genome of *w*Mol was released (Pinto et al., 2013), providing a unique opportunity to compare and contrast it with the previously sequenced and closely related *w*Pel genome (Salzberg et al., 2008). Because of the intracellular nature of *Wolbachia* and its dependence on its host for survival, these organisms bear reduced genomes. The genome of *w*Pel is 1.48 Mb long and encodes 1275 putative proteins. Similarly, *w*Mol has a genome of 1.34 Mb long encoding 1,191. The two reduced genomes are small enough to compare and sort through each gene individually. Logically, genomic differences between these two incompatible strains of *Wolbachia* should highlight the specific toxin-antidote system genes of CI.

Currently, all knowledge about the molecular mechanism of CI is limited to post-fertilization effects occurring on host paternal chromatin such as delays in cell cycle and chromatin condensation (Ryan and Saul, 1968; Reed and Werren, 1995; Callaini et al., 1997), chromosomal breakage or an inability to separate at anaphase (Callaini et al. 1997), and improper deposition of maternal histones (Landmann et al., 2009). Aside from these observations, the biochemistry of CI remains unexplained. Multiple CI effectors have recently been proposed including: ankyrin genes (Iturbe-Ormaetxe et al., 2005), the WO PHAGE (Serbus et al., 2008), the *Wolbachia* DNA Binding HU protein (Beckmann et al. 2012), a *Wolbachia* encoded transcription factor, *w*trM (Pinto et al., 2013), a histone chaperone HIRA (Zheng et al., 2011) and a hypothetical protein WPIP0282 (Beckmann

and Fallon 2013). Of these proteins, WPIP0282 is the only protein shown to be present at the right place and right time to induce CI. Specifically, we recently detected WPIP0282 within mature mosquito sperm derived from mosquito spermathecae as well as showed that *w*Pip_0282 is transcribed as a two gene operon with its partner *w*Pip_0283 (Fig 1A). This operon is the most scientifically supported candidate for CI induction/rescue not only because WPIP0282 was found on the sperm, but also because it is uniquely present in strains of *Wolbachia* that are known to induce CI and absent in those that don't. The copy number of operon paralogs within *Wolbachia* genomes correlates with CI intensity and *w*Pip_0283 encodes a putatively horizontally transferred eukaryotic Ulp1 SUMO protease domain. SUMO protease-like proteins have no known native function within prokaryotes except as host effectors in infectious bacteria; bacteria do not use ubiquitin based systems (Hotson et al., 2003; Misaghi et al., 2006). My hypothesis for CI induction involves expression of this two gene operon as a toxin-antidote system where WPIP0282 induces CI epigenetically in sperm chromatin and WPIP0283 rescues CI via the ubiquitin pathways in the zygote.

Results and Discussion:

Genomic comparison of wPip(Pel) with wPip(Mol). To test this hypothesis I compared the genomes of *w*Pel and *w*Mol with the specific intent of finding differences within the *w*Pip_0282-0283 operon reasoning that genomic differences between the two strains would account for the incompatibility between them. Upon genomic comparison we noted that the genomes were extremely similar with 816 proteins being exactly the

same, roughly 69% of the total encoded ORFS. In contrast to a previous report (Pinto et al., 2013), we noted that the rest of the ORFS encoded proteins that were different in some way between the two strains. In total I produced a subset list of 448 diverging proteins which logically must include the CI system. Most of these proteins are extremely similar and often encode only one amino acid difference. However, all these proteins were included for the sake of completeness, no matter how small the divergence. I then classified these 448 proteins according to their putative functions in order to further hone in on potential CI inducers.

The majority of diverging orthologs (26%) were viral/phage associated genes, 117 in total, which is not surprising because the WO phage is known to be particularly active and divergent within the *Culex* strains of *Wolbachia*. However, I believe the WO phage to be entirely independent of CI because extensive studies in *Culex* have shown no correlation of CI to the phage, and CI induced by *Wolbachia* occurs even in strains without phage (Serbus et al., 2008). Therefore it is unlikely that divergences in these phage proteins account for the incompatibility between *w*Mol and *w*Pel.

The second largest grouping of divergent proteins (18%) was mobile element genes such as transposases. I also considered these genes unlikely to be involved in CI (aside from being a catalyst for genomic divergence) because numerous bacteria infecting the reproductive tissues of arthropods have divergent mobile elements but are not known to induce reproductive phenotypes like CI. For instance *Rickettsia peacockii* and *Rickettsia rickettsii* have multiple differences with respect to transposable element genes (Felsheim et al., 2009), yet the ticks they infect (*Dermacenter andersoni)* have not been

noted to be incompatible due to these different endosymbionts. Therefore, the subset of potential CI inducers can be reduced to a list of 250 proteins by subtracting the divergent proteins belonging to the phage and mobile element groupings. One is then left with a short list of 250 proteins most likely to account for incompatibilities between the two strains.

Among these proteins most are nearly identical. Only a small proportion of the 250 have diverged to the extent that they share less than 75% homology and/or are truncated and missing greater than 10% of the protein. Upon inspection of these most divergent orthologs I immediately noted, as I had predicted, the putative toxin-antidote operon proteins WPIP0282 and WPIP0283 were among the most severely fragmented and divergent proteins in *w*Mol with respect to *w*Pel. The orthologous *w*Pip_0282-0283 operon in *w*Mol had been severely scrambled and scattered throughout a small area within the *w*Mol genome (Fig 1B). What was extremely unusual was that the orthologous operon was fragmented into six different pieces, implying at least 5 different genomic rearrangements of this particular operon. In fact, between *w*Mol and *w*Pel the relatively small genomic region of this operon contains the densest number of rearrangements within the entire genome. I believe that the scrambling of the operon in *w*Mol is strong evidence supporting the hypothesis that it is involved in CI; precisely because the two strains, *w*Mol and *w*Pel, are incompatible and yet only have a small subset of genes

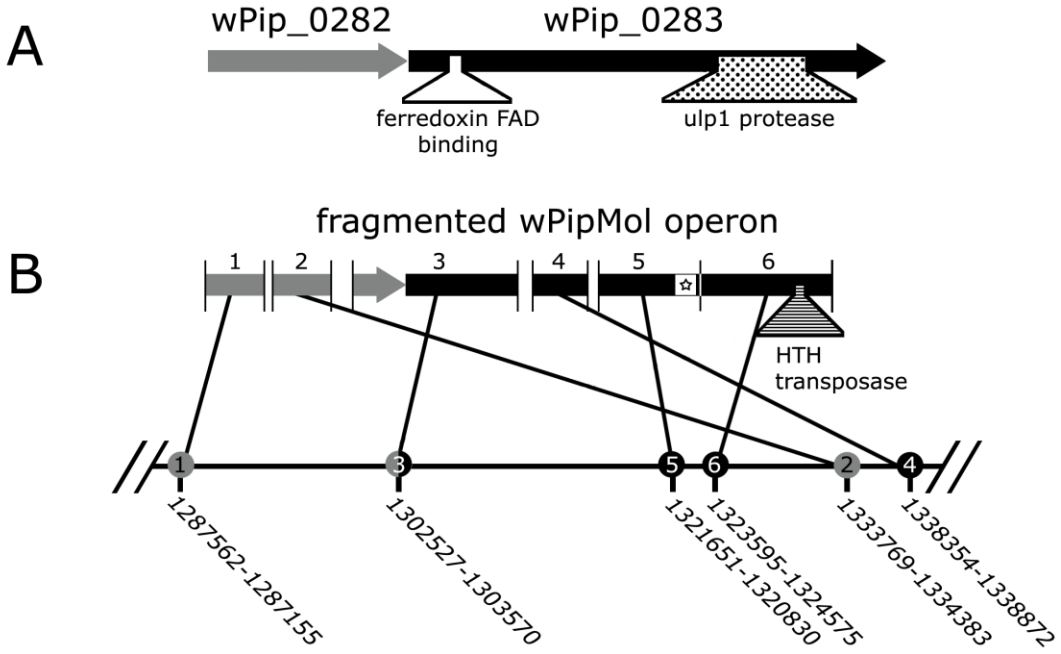

A
wPip_0282
wPip_0283
ferredoxin FAD binding
ulp1 protease
B
fragmented wPipMol operon
1
2
3
4
5
6
HTH transposase
1
3
5
6
2
4
1287562-1287155
1302527-1303570
1321651-1320830
1323595-1324575
1333769-1334383
1338354-1338872

Figure 1. Organization of the *w*Pip_0282-0283 operon. **A.** Intact operon from the genome of *w*PipPel. WPIP0282 shows no identifiable protein domains. WPIP0283 contains an FNR-like ferredoxin reductase FAD binding domain (superfamily cl06868) and a Ulp1 protease (superfamily cl177784). **B.** Fragmented pieces of the same operon in *w*PipMol and their genomic locations. Un-annotated ORFs are predicted for fragments 1, 3, and 5. The star in fragment 5 represents a non-homologous C-terminal addition with predicted (PSI-BLAST) homology to the catalytic domain of C-19 ubiquitinyl hydrolazes. Fragment 6 contains a helix-turn-helix DNA binding domain from a transposase which likely broke apart the operon. Fragment six is also missing the characteristic Ulp1 protease present in *w*PipPel.

Fragment	Homolog in wPipPel	Identities	Homology	ORF Prediction	Direction	start (5')	end (3')
1	0282	118/119	99%	Yes	<	1287562	1287155
2	*0282	185/205	90%	No	>	1333769	1334383
3	0282	85/91	93%	No	>	1302527	1303018
3	0283	215/239	90%	Yes	>	1302852	1303570
4	0283	141/172	82%	No	>	1338354	1338872
5	0283	244/244	100%	Yes	<	1321651	1320830
6	0283	139/272	51%	No	<	1323595	1324575
NA	1291	284/284	100%	Yes	<	1168537	1166977

Table 1. Orthologous wPip(Mol) operon fragments.

which are largely divergent between them of which the 0282/0283 operon is one of the most severely changed.

Interestingly, some fragments of the operon are well conserved and others not (Table 1), as well as some domains being present and others absent between the two strains (Fig 1B). We note that the fragments that are well conserved contain large un-annotated open reading frames such as fragment 1, 3, and 5 (Fig. 1). These fragments might still be expressed and may or may not encode proteins with increased/reduced functionality and/or changed functionality and abundance. Such a pattern would be consistent with the timing model of CI, which suggests that the CI toxin skews cell cycle timing in a way that causes the paternal pronucleus to slow down and become out of sync with the maternal pronucleus (Chapter 1). Furthermore, we note that *w*Mol fragment 6, which should putatively encode the Ulp1-like SUMO protease domain, in fact lacks this domain, which has been replaced by a fragment of a helix-turn-helix DNA binding domain from a transposase (Fig 1B). This transposase is possibly a remnant of the original transposase that scrambled the operon.

A Final Hypothesis for the Molecular Mechanism of CI. I previously suggested that in *w*Pel the Ulp1 SUMO protease domain was possibly involved with rescue of CI via ubiquitin pathways based upon it being encoded in an operon with the *Wolbachia* sperm factor *w*Pip_0282 and upon the assumption that it was horizontally transferred from an ancient insect, possibly a *Dipteran* and therefore is likely to interact with *Dipteran* proteins (Beckmann and Fallon 2013). We hypothesized that this protein was horizontally transferred based upon BLAST analysis, which indicated that the Ulp1 SUMO protease domain was most closely related to eukaryotic homologs within *Drosophila* (Beckmann and Fallon 2013; Chapter 4). It is also located within region of the *Wolbachia* genome known to be extremely mobile, being surrounded by pieces of the WO phage. The actual operon is immediately surrounded on the 5' and 3' ends by mobile DNA such as transposases, group-II reverse transcriptases, recombinases and integrases (Beckmann and Fallon 2013).

Because of the unique difference between *w*Pel and *wMol*, with respect to the Ulp1 SUMO protease domain protein, as well as BLAST analysis pointing to its eukaryotic origins, and its operon linkage to the spermatheca associated protein WPIP0282, I propose a ubiquitin/SUMO mediated hypothesis for CI (Figure 2). A *Wolbachia* CI induction/rescue system must intricately interact with or exert effects on insect chromatin because defects in this very system are the unanimous cytological observations in all CI susceptible insect crosses (Chapter 1.) To that point, ubiquitylation, SUMOylation and de-SUMOylation have already been shown to be key regulators of chromatin dynamics, mitotic progression, and cell cycle checkpoints acting via post-

translational modification of proteins such as, Topoisomerase-II, SMC proteins, and the Anaphase Promoting Complex, all of which are key regulators of anaphase chromatin structure and dynamics (Hickey et al., 2012; Strunnikov et al., 2001; Gutierrez and Ronai, 2006). Specifically, Ulp1 proteins have already been shown to localize in the eukaryotic nucleus where they de-SUMOylate substrates (Li and Hochstrasser, 2003; Felberbaum and Hochstrasser, 2008). I envision that the *w*Pip_0282-0283 operon behaves in a way similar to a toxin-antidote operon where WPIP0282 induces an effect upon the host ubiquitin sperm pathway thereby effecting paternal chromosome condensation at anaphase within the embryo and freezing cell cycle checkpoints prior to the completion of anaphase. Subsequently, these effects can be rescued in the egg by overexpressing the *Wolbachia* derived Ulp1 proteins WPIP0283 and WPIP1291.

To summarize, evidence in support of this model includes the following observations: 1) WPIP0282 was detected within sperm derived from insect spermathecae; 2) *w*Pip_0282 is contained within a two gene operon, also encoding the Ulp1 SUMO protease domain; 3) this operon is likely of eukaryotic origins and therefore adapted to uniquely manipulate host machinery; 4) the operon is only present within strains of CI-inducing *Wolbachia*; 5) there are major genomic differences with respect to the operon's structure and sequence between the similar but bi-directionally incompatible *Wolbachia* strains *w*Pel and *w*Mol. Finally, the operon has duplicated and diverged multiple times in multiple strains of *Wolbachia*, accounting for the multiple phenotypes, incompatibilities, and CI-strengths observed within insects (Beckmann and Fallon 2013).

In my postdoctoral work in the Hochstrasser lab, I will continue to enzymatically test these two proteins until I have verified or disproven their abilities to induce and rescue CI like effects. However, in a way, my model is already supported by mutagenesis in eukaryotic cells. Yeast Ulp mutants exhibit prolonged metaphase and have defects in centromeric cohesion as well as an inability to escape mitotic checkpoints (Bachant et al. 2002; Felberbaum and Hochstrasser, 2008) which is precisely the phenotype exhibited in CI crosses and outlined previously in the timing model of CI (Chapter 1). Furthermore, consistent with an involvement of Ulp1 in meiotic chromatin processes, transpositional mutagenesis studies within *Drosophila* targeting the native dmUlp1 created a strain (Ulp1^{G0026}) of sterile male flies (Peter et al. 2002). Therefore, because the phenotypes of Ulp mutants mimic CI cytologically, I predict that CI induction/recue is a result of modulation and interference with host Ulp proteins. I envision that WPIP0282 causes this interference in the insect sperm and WPIP0283 acts to rescue this interference by overexpression within *Wolbachia* infected eggs (Fig. 2).

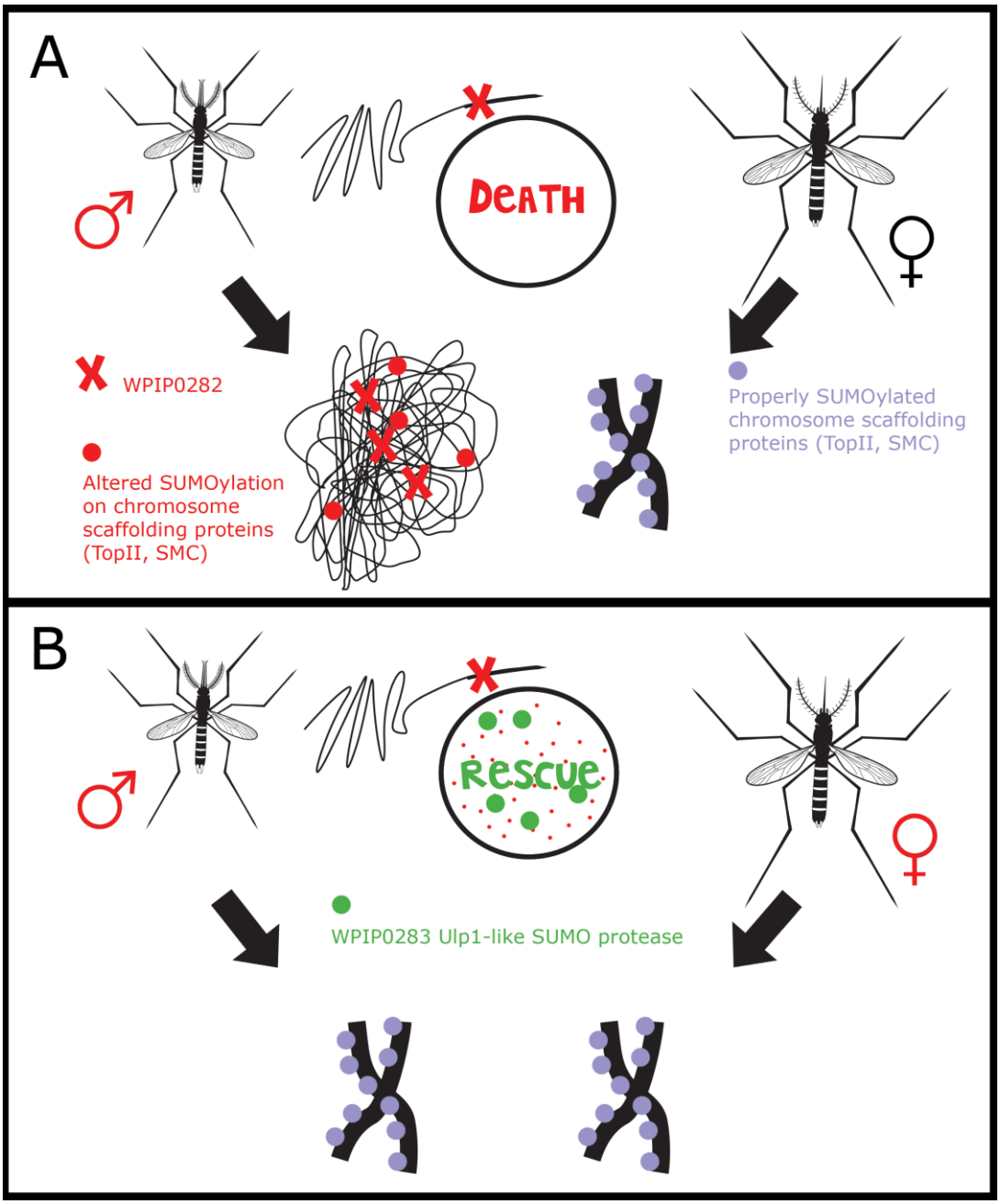

A
DEATH
WPIP0282
Altered SUMOylation on chromosome scaffolding proteins (TopII, SMC)
Properly SUMOylated chromosome scaffolding proteins (TopII, SMC)
B
Rescue
WPIP0283 Ulp1-like SUMO protease

Figure 2. The Ubiquitin/SUMO mediated hypothesis for the molecular mechanism of CI. **A.** WPIP0282 is expressed by *Wolbachia* in the testes and accumulates in mature sperm. WPIP0282 is proposed to cause an epigenetic alteration in the ubiquitination/SUMOylation status of chromatin scaffolding proteins Topoisomerase II, and Cohesin/Condensin complexes (SMC proteins) causing retarded condensation of paternal chromosomes during the first mitosis. Death is caused by shearing of paternal chromatin at anaphase. **B.** Expression of the putative rescue factor and SUMO protease, WPIP0283, in the egg, which is proposed to reverse these alterations and allow proper condensation of paternal chromosomes.

The results contained in this thesis demonstrate that among known potential CI inducing protein candidates, CI is most likely to involve the protein products of the *w*Pip0282-0283 operon. This research is extremely important because epidemiological applications of CI have been hindered by the fact that *Anopheles* mosquitoes (malarial vectors) and *Aedes aegypti* mosquitoes (dengue and yellow fever vectors) do not harbor natural *Wolbachia* infections, and *Anophelines* are recalcitrant to artificial infection by microinjection. Therefore, to use CI as a gene driving force within these vectors one must first identify the genetic mechanism underlying CI and directly clone these genes into transgenic disease vectors separately and independent of the *Wolbachia* organism. This will be the ultimate goal of the continuing work.

BIBLIOGRAPHY

Andersen J. D., Boylan K. L., Xue F. S., Anderson L. B., Witthuhn B. A., Markowski T.W., Higgins L., and Skubitz A. P. (2010). Identification of candidate biomarkers in ovarian cancer serum by depletion of highly abundant proteins and differential in-gel electrophoresis. Electrophoresis 31, 599-610.

Arez, A. P., D. Lopes, J. Pinto, A. S. Franco, G. Snounou, and V. E. do Rosario. (2000). Plasmodium sp.: optimal protocols for PCR detection of low parasite numbers from mosquito (*Anopheles sp.)* samples. Exp. Parasitol. 94:269-272.

Assier, E., Bouzinba-Segard, H., Stolzenberg, M., Stephens, R., Bardos, J., Freemont, P., Charron, D., Trowsdale, J., and Rich, T. (1999). Isolation, sequencing, and expression of RED, a novel human gene encoding an acidic-basic dipeptide repeat. Gene 230, 145-154.

Azam, T. A., Iwata, A., Nishimura, A., Ueda, S., and Ishihama, A. (1999). Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. J. Bacteriol. 181, 6361-6370.

Bachant, J., Alcasabas, A., Blat, Y., Kleckner, N., and Elledge S. J. (2002). The SUMO-1 isopeptidase Smt4 is linked to centromeric cohesion through SUMO-1 modificatino of DNA topoisomerase II.

Ballhorn R. (2007). The protamine family of sperm nuclear proteins. Genome Biol. 8, 227.

Barr, R. A., (1980). Cytoplasmic Incompatibility in Natural Populations of a Mosquito, *Culex pipiens* L. Nature 283, 71-72

Bazzocchi, C., Comazzi, S., Santoni, R., Bandi, C., Genchi, C., and Mortarino, M. (2007) *Wolbachia* surface protein (WSP) inhibits apoptosis in human neutrophils. Parasite Immunol. 29, 73-79.

Beckmann. J. F., and Fallon, A. M. (2012) Decapitation improves detection of *Wolbachia pipientis* (*Rickettsiales: Anaplasmataceae*) in *Culex pipiens* (Diptera: Culicidae) mosquitoes by the polymerase chain reaction. J. Med. Entomol. 49, 1103-1108.

Beckmann, J. F. and Fallon, A. M. (2013). Detection of the *Wolbachia* protein WPIP0282 in mosquito spermathecae: Implications for cytoplasmic incompatibility. *J. Ins. Bioc. Mol. Biol. 43(9), 867-78.*

PMCID: PMC3839623

Beckmann, J. F., Markowski, T. W., Witthuhn, B. A., and Fallon, A. M. (2013) Detection of the *Wolbachia*-encoded DNA binding protein, HU beta, in mosquito gonads. Insect Biochem. Mol. Biol. 43, 272-279.

Benjamin, B., Yves, B., and Corinne, A. (2007) Assembly of the Tc1 and *mariner* transposition initiation complexes depends on the origins of their transposase DNA binding domains. Genetica, 130, 105-120.

Berticat, C., F. Rousset, M. Raymond, A. Berthomieu, and M. Weill. (2002). High *Wolbachia* density in insecticide-resistant mosquitoes. Proc R. Soc. Lond. B 269: 1413-1416.

Bianchi, M. E. (1994). Prokaryotic HU and eukaryotic HMB1: a kinked relationship. Mol. Microbiol. 14, 1-5.

Bordenstein, S. R. and Reznikoff, W. S. (2005) Mobile DNA in obligate intracellular bacteria. Nature Rev. Microbiol. 3, 688-699.

Bourtzis, K., Nirgianaki, A., Markakis, G., and Savakis, C. (1996) *Wolbachia* infection and cytoplasmic incompatibility in *Drosophila* species. Genetics 144, 1063-1073.

Braig, H. R., Zhou, W., Dobson, S. L., and O'Neill, S. L. (1998). Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. J. Bacteriol. 180, 2373-2378.

Brattig, N. W., Bazzochi, C., Kirschning, C. J., Reiling, N., Buttner, D. W., Ceciliani, F., Geisinger, F., Hochrein, H., Ernst, M., Wagner, H., Bandi, C., and Hoerauf, A. (2004) The major surface protein of *Wolbachia* endosymbionts in filarial nematodes elicits immune responses through TLR2 and TLR4. J Immunol. 173, 437-445.

Breeuwer, J. A. J., and Werren J. H. (1990). Microorganisms associated with chromosome destruction and reproductive isolation between two insect species. Nature 346, 558-560.

Byrne, K. and Nichols R. A. (1999). Culex pipiens in London Underground tunnels: differentiation between surface and subterranean populations. Heredity 82, 7-15

Callaini, G., Dallai, R., and Giovanna Riparbelli, M. (1997). *Wolbachia*-induced delay of paternal chromatin condensation does not prevent maternal chromosomes from entering anaphase in incompatible crosses of *Drosophila simulans*. J. Cell Sci. 110, 271-280.

Caspari, E., and Watson G., S. (1959). On the Evolutionary Importance of Cytoplasmic Sterility in Mosquitoes. Evolution 13, 568-570.

Chen, Y. P., J. S. Pettis, A. Collins, and M. F. Feldlaufer. (2006). Prevalence and transmission of honeybee viruses. Appl. Environ. Microbiol. 72: 606-611.

Clark, M. E., Bailey-Jourdain, C., Ferree, P. M., England, S. J., Sullivan, W., Windsor, D. M., and Werren, J. H. (2008). *Wolbachia* modification of sperm does not always require residence within developing sperm. Heredity 101, 420-428.

Clark, M. E., Veneti, Z., Bourtzis, K., and Karr, T. (2002). The distribution and proliferation of the intracellular bacteria *Wolbachia* during spermatogenesis in *Drosophila*. Mech. of Dev. 111, 3-15.

Clark, M. E., Veneti, Z., Bourtzis, K., and Karr, T. (2003). *Wolbachia* distribution and cytoplasmic incompatibility during sperm development: the cyst as the basic cellular unit of CI expression. Mech. of Dev. 120, 185-198.

Cornelis, G. R. and Denecker, G. (2001) *Yersinia* lead SUMO attack. Nature Med. 7, 21-23.

Cowdry, E., V. (1923). The Distribution of *Rickettsia* in the Tissues of Insects and Arachnids. Journ. Exp. Med., xxxvii, 431-456.

Cubenas-Potts, C. and Matunis, M. J. (2013) SUMO: a multifaceted modifier of chromatin structure and function. Dev. Cell 24, 1-12.

Darby, A. C. et al. (2012). Analysis of gene expression from the *Wolbachia* genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. Genome Res. [*Epub ahead of print*]

Deshaies, R. J. (1999) SCF and cullin/RING H2-based ubiquitin ligases. Annu. Rev. Cell Dev. Biol. 15, 435-467.

Dobson, S. L., K. Bourtzis, H. R. Braig, B. F. Jones, W. Zhou, F. Rousset, and S. L. O'Neill. (1999). *Wolbachia* infections are distributed throughout insect somatic and germ line tissues. Insect Biochem. Mol. Biol. 29: 153-160.

Dorman, C. J., (2009). Nucleoid-associated proteins and bacterial physiology. Adv. Appl. Microbiol. 67, 47-64.

Duron, O., J. Lagnel, M. Raymond, K. Bourtzis, P. Fort, and M. Weill. (2005).

Transposable element polymorphism of *Wolbachia* in the mosquito *Culex pipiens*: evidence of genetic diversity, superinfection and recombination. Mol. Ecol. 14: 1561-1573.

Echaubard, P., O. Duron, P. Agnew, C. Sidobre, V. Noel, M. Will, and Y. Michalakis. (2010). Rapid evolution of *Wolbachia* density in insecticide resistant *Culex pipiens*. Heredity 104: 15-19.

Edgar, R. C. (2004), MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32, 1792-1797.

Fallon, A. M. (2008). Cytological properties of an *Aedes albopictus* mosquito cell line infected with *Wolbachia* strain *w*AlbB. In vitro Cell Dev. Biol. – Animal 44: 154-161.

Fallon, A. M., Baldridge, G.D. Higgins, LA, and Witthuhn, BA. (2013) *Wolbachia* from the planthopper *Laodelphax striatellus* establishes a robust, persistent, streptomycin-resistant infection in clonal mosquito cells. In Vitro Cell. Dev. Biol. Anim. 49, 66-73.

Fallon, A. M., and Witthuhn, B. A. (2009). Proteasome activity in a naïve mosquito cell line infected with *Wolbachia pipientis w*AlbB. In Vitro Cell. Dev. Biol. – Animal 45, 460-466.

Felberbaum, R. and Hochstrasser, M. (2008). Ulp2 and the DNA damage response: Desumoylation enables safe passage through mitosis. Cell Cycle 7, 52-56

Felsheim, R. F., Kurtti, T. J., and Munderloh, U. G. (2009). Genome sequence of the endosymbiont *Rickettsia peacockii* and comparison with virulent *Rickettsia rickettsia*: identification of virulence factors. PLoS One, 4, e8361.

Ferree, P. M., Frydman, H. M., Li, J. M., Cao, J., Wieschaus, E., and Sullivan, W. (2005). *Wolbachia* utilizes host microtubules and dynein for anterior localization in the *Drosophila* oocyte. PLoS Pathog. 1, e14.

Fine, P., E., M. (1978). On the Dynamics of Symbiote-Dependent Cytoplasmic Incompatibility in Culicine Mosquitoes. J. Inv. Pathol. 30, 10-18

Frentiu, F. D., Robinson, J., Young, P. R., McGraw, E. A., and O'Neill, S. L. (2010). *Wolbachia*-mediated resistance to dengue virus infection and death at the cellular level. *PloS* One 5, e13398.

Friedman, D. I. (1988). Integration host factor: a protein for all reasons. Cell 55, 545-554.

Gutierrez, G. J. and Ronai, Z. (2006). Ubiquitin and USMO systems in the regulation of mitotic checkpoints. Trends Biochem Sci. 31, 324-332.

Harden, V. A. (1990). Rocky Mountain Spotted Fever: History of a Twentieth-Century Disease, John's Hopkins University Press Baltimore, Maryland 21211

Hennig, W. (2003). Chromosomal proteins in the spermatogenesis of *Drosophila.* Chromosoma 111, 489-494.

Hertig, M. and Wolbach, S. B. (1924). Studies on *Rickettsia*-like Micro-organisms in Insects. J. med. Res. 44, 329-74

Hertig, M. (1936). The *Rickettsia*, *Wolbachia pipientis* (Gen. et sp.n.) and Associated Inclusions of the Mosquito, *Culex pipiens*. Parasitol. 28, 453-486.

Hickey, C. H., Wilson, N. R., and Hochstrasser, M. (2012). Function and regulation of SUMO proteases. Nature Rev. Mol. Biol. 13, 755-766

Hoffmann, A.A., B.L. Montgomery, J. Popovici, I. Iturbe-Ormaetxe, P.H. Johnson, F. Muzzi, M. Greenfield, M. Durkan, Y.S. Leong, Y. Dong, H. Cook, J. Axford, A.G. Callahan, N. Kenny, C. Omodei, E.A. McGraw, P.A. Ryan, S.A. Ritchie, M. Turelli, and S.L. O'Neill. (2011). Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. Nature 476, 454-457.

Hoffmann, A., A., Turelli, M., and Simmons, G., M. (1986). Unidirectional Incompatibility between Populations *of Drosophila simulans*. Evolution 40, 692-701

Hoffmann, A. A., Turelli, M., and Harshman, L. G. (1990). Factors Affecting the Distribution of Cytoplasmic Incompatibility in *Drosophila simulans*. Genetics 126, 933-948

Hotson, A., Chosed, R., Shu, H., Orth, K. and Mudgett, M. B. (2003) Xanthomonas type III effector XopD targets SUMO-conjugated proteins in *planta*. Mol. Microbiol. 50, 377-389. PMID: 14617166

Hughes, G. L., R. Koga, P. Xue, T. Fukatsu, and J.L. Rasgon. (2011). *Wolbachia* infections are virulent and inhibit the human malaria parasite *Plasmodium falciparum* in *Anopheles gambiae*. PLoS Pathog 7: e1002043.

Ishmael, N., Dunning Hotopp, J. C., loannidis, P., Biber, S., Sakamoto, J., Siozios, S.,

Nene, V., Werren, J., Bourtzis, K., Bordenstein, S. R., and Tettelin, H. (2009) Extensive genomic diversity of closely related *Wolbachia* strains. Microbiology 155, 2211-2222.

Itah, R., I. Gitelman, and C. Davis. (2004). A replacement for methoxyflurane (Metofane) in open-circuit anaesthesia. Laboratory Animals 38, 280-285.

Iturbe-Ormaetxe, I., Burke, G. R., Riegler, M., and O'Neill, S. L. (2005). Distribution, expression, and motif variability of ankyrin domain genes in *Wolbachia pipientis*. J. Bacteriol. 187, 5136-45.

Jaffe, A., Vinella, D., and D'Ari, R. (1997). The *Escherichia coli* histone-like protein HU affects DNA initiation, chromosome partitioning via *Muk*B, and cell division via *Min*CDE. J. Bacteriol. 179, 3494-3499.

Jeyaprakash, A. and M. A. Hoy. (2000). Long PCR improves *Wolbachia* DNA amplification: *wsp* sequences found in 76% of sixty-three arthropod species. Insect Mol. Biol. 9: 393-405.

Jost, E. (1970). Genetische Untersuchungen zur Inkompatibilitat im Culex-pipiens-Komplex. Theoret. Appl. Gen. 40, 251-256

Karr, T. L., Yang, W., and Feder, M. E. (1998). Overcoming cytoplasmic incompatibility in *Drosophila.* Proc. R. Soc. Lond. B. 265, 391-395.

Keller A., Nesvizhskii A. I., Kolker E., Aebersold R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. 74, 5383-5392.

Kiger, A. A., White-Cooper, H., and Fuller, M. T. (2000). Somatic support cells restrict germline stem cell self-renewal and promote differentiation. Nature 407, 750-754.

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., and Parkhill, J. (2008) Genome evolution of *Wolbachia* strain *w*Pip from the *Culex pipiens* group. Mol. Biol. Evol. 25, 1877-1887.

Knizewski, L., Kinch, L. N., Grishin, N. V., Rychlewski, L., and Ginalski, K. (2007) Realm of PD-(D/E)XK nuclease superfamily revisited: detection of novel families with modified transitive meta profile searches. BMC Struc. Biol. 7:40, doi:10.1186/1472-6807-7-40.

Kriesner, P., Hoffmann, A. A., Lee, S. F., Turelli, M., and Weeks, A. R. (2013.) Rapid

Sequential Spread of Two Wolbachia Variants in Drosophila simulans. PLOS pathogens 9, e1003607

Kumar. S., Sardesai, A. A., Basu, D., Muniyappa, K., and Hasnain, S. E. (2010). DNA clasping by mycobacterial HU: the C-terminal region of *hup*B mediates increased specificity of DNA binding. PLoS One 5, e12551.

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

Laven, H. (1953.) Rezprok unterschiedliche Kreuzbarkeit von Stechnucken (*Culicidae*) und ihre Deutung als plasmatische Vererbung, Z. Vererbungsl., 85, 118-136

Laven, H. (1957) Vererbung durch Kerngene und das Problem der ausserkaryotischen Vererbung bei *Culex pipiens*, II. Ausserkaryotische Vererbung, Z. Vererbungsl., 88, 478-516

Laven, H. (1967a). A Possible Model for Speciation by Cytoplasmic Isolation in the *Culex pipiens* Complex. Bull. Org. mond. Sante. 37, 263-266

Laven, H. (1967b). Chapter 7: Speciation and Evolution in *Culex pipiens*. In genetics of Distinct Vectors of Disease. Ed. J. Wright. Amsterdam Elselvier 251-275.

Laven, H. (1967c). Eradication of Culex pipiens fatigans through Cytoplasmic Incompatibility. Nature 216, 383-384

Lee, M., Bakir, A. A., Nguyen, K. N., Bachant, J. (2011) The SUMO isopeptidase Ulp2p is required to prevent recombination-induced chromosome segregation lethality following DNA replication stress. PLos Genetics 7, e1001355.

Lee, N., Francklyn, C., and Hamilton, P. E., (1987). Arabinose-induced binding of AraC protein to araI2 activates the araBAD operon promoter. Proc. Natl. Acad. Sci. 84, 8814-8818

Li, S. and Hochstrasser, M. (1999) A new protease required for cell-cycle progression in yeast. Nature 398, 246-251.

Li, S. and Hochstrasser, M. (2003). The Ulp1 SUMO isopeptidase: distinct domains required for viability, nuclear envelope localization, and substrate specificity. J. Cell Biol. 160, 1069-1081

Livak, K. J. 1984. Organization and mapping of a sequence on the *Drosophila*

melanogaster X and Y chromosomes that is transcribed during spermatogenesis. Genetics 107: 611-634.

Lobley, A., Sadowski, M. I., and Jones D. T. (2009) pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. Structural Bioinformatics 25, 1761-1767.

Lorenzen, M. D., Gnirke, A., Margolis, J., Garnes, J., Campbell, M., Stuart, J. J., Aggarwal, R., Richards, S., Park, Y., and Beeman, R. W. (2008) The maternal-effect, selfish genetic element Medea is associated with a composite Tc1 transposon. Proc. Natl. Acad. Sci. USA 105, 10085-10089.

Magnin, M., Pasteur, N., and Raymond, M. (1987) Multiple incompatibilities within populations of *Culex pipiens L.* in southern France. Genetica 74, 125-130.

Marshall, J. f., 1938. The British Mosquitoes, London, (1938).

McAleenan, A., Cordon-Preciado, V., Clemente-Blanco, A., Liu, I., Sen, N., Leonard, J., Jarmuz, A., and Aragon, L. (2012) SUMOylation of the α-kleisin subunit of cohesion is required for DNA damage-induced cohesion. Curr. Biol. 22, 1-12.

McMeniman, C. J., Lane, R. V., Cass, B.N., Fong, A. W. C., Sidhu, M., Wang, Y., and O'Neill, S. L. (2009). Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti*. Science 323, 141-144.

Misaghi, S., Balsara, Z. R., Catic, A., Spooner, E., Ploegh, H. L., and Starnbach, M. N. (2006) *Chlamydia trachomatis*-derived de-ubiquitinating enzymes in mammalian cells during infection. Mol. Microbiol. 61, 142-150. PMID: 16824101

Moreira, L.A., I. Iturbe-Ormaetxe, J.A. Jeffery, G. Lu, A.T. Pyke, L.M. Hedges, B.C. Rocha, S. Mendelin, A. Day, M. Riegler, L.E. Hugo, K.N. Johnson, B.H. Kay, E.A. McGraw, A.F. van den Hurk, P.A. Ryan, and S.L. O'Neill. 2009. A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, Chikungunya, and *Plasmodium*. Cell 139, 1268-1278.

Nash, H. A., and Robertson, C. A. (1981). Purification and properties of the *Escherichia coli* protein factor required for λ integrative recombination. J. Biol. Chem. 256, 9246-9253.

Nasmyth, K., and Haering, C. H. (2005). The structure and function of SMC and kleisin complexes. Annu. Rev. Biochem. 74, 595-648.

Ner, S. S., and Travers, A. A. (1994). HMG-D, the *Drosophila melanogaster*

homologue of HMG 1 protein, is associated with early embryonic chromatin in the absence of histone H1. EMBO J. 13, 1817-1822.

Nesvizhskii, A. I., Keller, A., Kolker E., Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 75, 4646-4658.

Noda, H., Y. Koizumi, Q. Zhang, and K. Deng. 2001. Infection density of *Wolbachia* and incompatibility level in two planthopper species, *Laodelphax striatellus* and *Sogatella furcifera*. Insect Biochem. Mol. Biol. 31: 727-737.

Noda, H., Miyoshi, T., and Koizumi, Y. (2002). In vitro cultivation of *Wolbachia* in insect and mammalian cell lines. In Vitro Cell. Dev. Biol. – Animal 38, 423-427.

Oberto, J., Drlica, K., and Rouviere-Yaniv, (1994). Histones, HMG, HU, IHF: Meme combat. Biochimie 76, 901-908.

O'neill, S. L., and Karr, T. L. 1990. Bidirectional Incompatibility Between Conspecific Populations of *Drosophila simulans*. Nature, 348, 178-180

O'Neill, S. L., M. M. Pettigrew, S. P. Sinkins, H. R. Braig, T. G. Andreadis, and R. B. Tesh. 1997. *In vitro* cultivation of *Wolbachia pipientis* in an *Aedes albopictus* cell line. Insect Mol. Biol. 6: 33-39.

O'Neill, S., R. Giordano, A. M. E. Colbert, T. L. Karr, and H. M. Robertson. 1992. 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. Proc. Natl. Acad. Sci. USA 89: 2699-2702.

Orth, K., Xu, Z., Mudgett, M. B., Bao, Z. Q., Palmer, L. E., Bliska, J. B., Mangel, W. F., Staskawicz, B., and Dixon, J. E. (2000) Disruption of signaling by *Yersinia* effector YopJ, a ubiquitin-like protein protease. Science 290, 1594-1597

Peter et al. (2002). Mapping and identification of essential gene functions on the X chromosome of *Drosophila*. EMBO Rep. 3, 34-38

Pinson, V., Takahashi, M., and Rouviere-Yaniv, J. (1999). Differential binding of the *Escherichia coli* HU, homodimeric forms and heterodimeric form to linear, gapped and cruciform DNA. J. Mol. Biol. 287, 485-497.

Pinto, S. B., Stainton, K., Harris, S., Kambris, Z., Sutton, E. R., Bonsall, M. B., Parkhill, J., and Sinkins, S. P. (2013). Transcriptional regulation of Culex pipiens mosquitoes by Wolbachia influences cytoplasmic incompatibility. PLos Pathog. 9, doi: 10.1371/journal.ppat. 1003647

Presgraves, D.C. (2000). A genetic test of the mechanism of *Wolbachia*-induced cytoplasmic incompatibility in *Drosophila.* Genetics 154: 771-776.

Potaro, J.K. and A. R. Barr. 1975. "Curing" *Wolbachia* infections in *Culex pipiens*. J. Medical Entomol. 12, 265.

Rankin, D. J., Rocha, E. P. C., and Brown, S. P. (2011) What traits are carried on mobile genetic elements and why? Heredity 106, 1-10.

Rances, E., Voronin, D., Tran-Van, V., and Mavingui, P. (2008). Genetic and functional characterization of the type IV secretion system in *Wolbachia.* J. Bacteriol. 190, 5020-5030.

Rasgon, J. L., Ren, X., and Petridis, M. (2006). Can *Anopheles gambiae* be infected with *Wolbachia pipientis*? Insights from an in vitro system. Appl. and Environ. Microb. 72, 7718-7722.

Realini, C., Rogers, S. W., and Rechsteiner, M. (1994). Keke motifs, proposed roles in protein-protein association and presentation of peptides by MHC class I receptors. FEBS Letters. 348, 109-113.

Reed, K. M. and Werren, J. H. 1995. Induction of Paternal Genome Loss by the Paternal-Sex-Ratio Chromosome and Cytoplasmic Incompatibility bacteria (*Wolbachia*): A Comparative study of early Embryonic Events. Mol. Reproduc. Dev. 40, 408-418

Richardson, P. M., Holmes, P. W., and Saul II, G. B. 1987. The Effect of Tetracycline on Nonreciprocal Cross Incompatibility in *Mormoniella* [= *Nasonia*] *vitripennis*. J. Inv. Pathol. 50, 176-183

Robich, R. M. and D. L. Denlinger. 2005. Diapause in the mosquito *Culex pipiens* evokes a metabolic switch from blood feeding to sugar gluttony. Proc Natl Acad Sci USA 102: 15912-15917.

da Rocha-Lima, H. 1916a. Beobachtungen bei Flecktyphuslausen. Arch. F. Schiffs- u. Trop. Hyg., xxi, 17-31

da Rocha-Lima, H. 1916b. Zur Aetiologie des Fleckfiebers. Berlin, kiln. Wochenschr. 567-569

Ryan, V. T., Grimwade, J. E., Nievera, C. J., and Leonard, A. C., (2002). IHF and HU stimulate assembly of pre-replication complexes at *Escherichia coli ori*C by two different mechanisms. Mol. Microbiol. 46, 113-124.

Ryan, S. L. and Saul II, G. B. 1968. Post-Fertilization Effect of Incompatibility Factors in

Mormoniella. Molec. Gen. Genetics 103, 29-26

Saida, F., Uzan, M., Odaert, B., and Bontems, F. 2006. Expression of highly Toxic Genes in E. coli: Special Strategies and Genetic Tools. Cur. Prot. Pep. Sci. 7, 47-56.

Saida F. 2007. Overview on the Expression of Toxic Gene Products in *Escherichia coli.* Cur. Protocol. Prot. Sci. 5.19

Salzberg, S. L., Puiu, D., Sommer, D. D., Nene, V., Lee, N. H. (2008) Genome sequence of the *Wolbachia* endosymbiont of *Culex quinquefasciatus* JHB. J. Bacteriol. 191, 1725.

Schriefer, M. E., J. B. Sacci, R. A. Wirtz, and A. F. Azad. 1991. Detection of polymerase chain reaction-amplified malarial DNA in infected blood and individual mosquitoes. Environ. Parasitol. 73: 311-316

Sasaki, T., Braig, H. R., and O'Neill, S. L. (1998). Analysis of *Wolbachia* protein synthesis in *Drosophila* in vivo. Insect Mol. Biol. 7, 101-105.

Serbus, L. R., C. Casper-Lindley, F. Landmann, and W. Sullivan. 2008. The genetics and cell biology of *Wolbachia*-Host Interactions. Annu. Rev. Genet. 42: 683-707.

Shiny, C., Krushna, N. S. A., Babu, S., Elango, S., Manokaran, G., and Narayanan, R. B. (2011) Recombinant *Wolbachia* heat shock protein 60 (HSP60) mediated immune responses in patients with lymphatic filariasis. Microbes Infect. Doi:10.1016/j.micinf.2011.07.004.

Siozios, S., Ioannidis, P., Klasson, L., Andersson, S. G. E., Braig, H. R., Bourtzis, K. (2013) The diversity and evolution of *Wolbachia* ankyrin repeat domain genes. PLOS One 8, e55390.

Sinkins, S. P. 2004. *Wolbachia* and cytoplasmic incompatibility in mosquitoes. Insect Biochem. Mol. Biol. 34: 723-729.

Strunnikov, A. V., Aravind, L., and Koonin, E. V. (2001). *Saccharomyces cerevisiae* SMT4 encodes an evolutionarily conserved protease with a role in chromosome condensation regulation. Genetics 158, 95-107.

Subbarao, S. K., Krishnamurthy, B. S., Curtis, C. F., Adak, T., and Chandrahas, R. K. 1977. Segregation of Cytoplasmic Incompatibility Properties in *Culex pipiens fatigans*. Genetics 87, 381-390

Swinger, K. K., and Rice, P. A. (2004). IHF and HU: flexible architects of bent DNA. Curr. Opin. Struc. Biol. 14, 28-35.

Taylor, M. J., Voronin, D., Johnston, K. L., and Ford, L. (2012) *Wolbachia* filarial interactions. Cell Microbiol. Doi:10.1111/cmi.12084

Tazuke, S. I., Schulz, C., Gilboa, L., Fogarty, M., Mahowald, A. P., Guichet, A., Ephrussi, A., Wood, C. G., Lehmann, R., and Fuller, M. T. (2002). A germline specific gap junction protein required for survival of differentiating early germ cells. Development 129, 2529-2539.

Teixeira, L., Ferreira, A., Ashburner, M. (2008) The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. PloS Biol. 6(12):e1000002.doi: 10.1371/journal.pbio.1000002.

Turelli, M. and Hoffmann, A., A. 1991. Rapid Spread of an Inherited incompatibility factor in California Drosophila. Nature 353, 440-442

Unckless, R. L., L. M. Boelio, J. K. Herren, and J. Jaenike. 2009. *Wolbachia* as populations within individual insects: causes and consequences of density variation in natural populations." Proc. R. Soc. B. 276: 2805-2811.

Vavre, F. and Charlat, S. (2012) Making (good) use of *Wolbachia*: what the models say. Curr. Opin. Microbiol. 15, 263-268.

Veneti, Z., Clark, M. E., Zabalou, S., Karr, T. L., Savakis, C., and Bourtzis K. (2003) Cytoplasmic incompatibility and sperm cyst infection in different *Drosophila-Wolbachia* associations. Genetics 164: 545-552.

Veneti Z., Clark, M. E., Karr, T. L., Savakis, C., and Bourtzis, K. (2004) Heads or Tails: Host-parasite interactions in the *Drosophila-Wolbachia* system. Appl. Environ. Microbiol. 70: 5366-5372.

Vergunst, A. C., van Lier, M. C. M., Dulk-Ras, A. D., Grosse Stuve, T. A., Ouwehand, A., and Hooykaas, P. J. J. (2005). Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of *Agrobacterium*. Proc. Natl. Acad. Sci. USA. 102, 832-837.

Vezzani D., M. Mesplet, and D. F. Eiras. 2011. PCR detection of *Dirofilaria immitis* in *Aedes aegypti* and *Culex pipiens* from urban temperate Argentina. Parasitol. Res. 108: 985-989

Wagar, E. A., and Stephens, R. S. (1988). Developmental-form-specific DNA-binding proteins in *Chlamydia* spp. Infect. Immun. 56, 1678-1684.

Wimmer, P., Schreiner, S., and Dobner, T. (2012) Human pathogens and the host cell

SUMOylation system. J. Virol. 86 642-654.

Worrall, A. F. and Connolly, B. A. 1990. The Chemical Synthesis of a Gene Coding for Bovine Pancreatic DNase I and Its Cloning and Expression in *Escherichia coli*. J. Biol. Chem. 265, 21889-21895

Xi, Z., C. C. H. Khoo, and S. L. Dobson. 2005. *Wolbachia* establishment and invasion in an *Aedes aegypti* laboratory population. Science 310: 326-328.

Yamaguchi Y., Park, J., and Inouye, M. (2011) Toxin-antitoxin systems in bacteria and archaea. Annu. Rev. Genet. 45, 61-79.

Yen, J. H., and Barr, R. A., 1971. New Hypothesis of the Cause of Cytoplasmic Incompatibility in *Culex pipiens* L. Nature 232, 657-658

Yen, J., H. and Barr, R., A. 1973. The Etiological Agent of Cytoplasmic Incompatibility in *Culex pipiens*. J. Inv. Pathol. 22, 242-250

Yue, C. and E. Genersch. 2005. RT-PCR analysis of deformed wing virus in honeybees (*Apis mellifera)* and mites (*Varroa destructor).* J. Gen. Virol. 86: 3419-3424.

Zdobnov E.M. and Apweiler R. (2001) InterProScan - an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17: 847-848.

Zheng, Y., Ren, P. P., Wang, J. L., and Wang, Y. F. (2011). *Wolbachia*-induced cytoplasmic incompatibility is associated with decreased Hira expression in male *Drosophila*. PloS One, 6, e19512

Zhou, W., F. Rousset, and S. L. O'Neill. 1998. Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. Proc. R. Soc. Lond. B 265: 509-515.

Zioni, N., V. Soroker, and N. Chejanovsky. 2011. Replication of *Varroa destructor* Virus 1 (VDV-1) and a *Varroa destructor* Virus 1 – Deformed Wing Virus Recombinant (VDV- 1-DWV) in the head of the honeybee. Virology 417: 106-112