Investigating the biochemical function of wVitA Type IV effectors, the Wolbachia of Nasonia vitripennis

Heidi Park
Wellesley College, hpark3@wellesley.edu

Follow this and additional works at: http://repository.wellesley.edu/thesiscollection

Recommended Citation
Park, Heidi, "Investigating the biochemical function of wVitA Type IV effectors, the Wolbachia of Nasonia vitripennis" (2012). Honors Thesis Collection. Paper 23.
Investigating the biochemical function of wVitA Type IV effectors, the Wolbachia of Nasonia vitripennis

Heidi Park

Advisors: Irene L.G. Newton (Indiana University) and Heather R. Mattila

Department of Biological Sciences
Wellesley College

This material is submitted as partial fulfillment of a B.A. degree with honors in Biological Sciences

©2012 by Heidi Park and Heather R. Mattila
Acknowledgements

I would like to thank Dr. Irene Newton for introducing me to scientific research and for her continued mentorship, even from Bloomington, IN.

I would also like to thank Dr. Heather Mattila for agreeing to serve as my on-campus advisor and making this thesis possible.

I would like to thank my committee members Dr. Mary Allen and Dr. Jennifer Hood-DeGrenier for their constant support and advice.

I would also like to thank the Biological Sciences Department Chair, Dr. John Cameron for making this thesis possible and providing administrative support.

I would also like to thank Dr. Jennifer Hood-DeGrenier, Dr. Yuichiro Suzuki, and Padma Kannabiran from the Biology Prep Room for generously allowing me to borrow laboratory supplies during the course of my thesis.

Lastly, I would like to thank my friends, especially Elaine Kim, Michelle Lee, Michelle Xu, and Luyang Jin for their love and support during this entire process.
Abstract

The obligately intracellular bacterium *Wolbachia pipiensis* is a mutualist of filarial nematodes and a reproductive parasite of insects. Currently, the molecular mechanisms the bacteria use to manipulate their eukaryotic hosts have yet to be elucidated. The *Wolbachia* genome encodes for a Type IV secretion system. It is possible that *Wolbachia* uses this secretion system to secrete proteins, called effectors, into the host intracellular environment to manipulate their hosts. This study looks specifically at wVitA, the *Wolbachia* strain found in the parasitic wasp, *Nasonia vitripennis*. As effectors tend to have 1) eukaryotic homology or 2) contain eukaryotic domains, the wVitA genome was queried using BLASTp (Genbank) and HMMer (Pfam). A total of 295 candidate *Wolbachia* eukaryotic-like genes (WEGs) were identified. wVitA effectors were identified using an assay developed by Lesser and Miller (2001) to identify Type IV effectors. Out of 32 tested proteins, 10 caused growth defects in wildtype yeast. Eight out of these 10 have a punctate cytoplasmic localization. It is hypothesized that these proteins may be involved in the inhibition of phagosome-lysosome fusion.
Introduction

Heinrich Anton de Bary defined symbiosis as “the living together of dissimilar organisms.” {Bary, 1879 #24} The word ‘symbiosis’ describes a continuum of relationships that encompasses mutualism (where both partners benefit), commensalism (where one partner benefits, but the other neither benefits nor is harmed), and parasitism (where one partner benefits, but the other is harmed).

The bacterial species Wolbachia pipiensis (class alpha-proteobacteria, order Rickettsiales) encompasses a group of obligately intracellular bacteria that can form a broad range of symbiotic relationships with their eukaryotic arthropod and nematode hosts. Wolbachia are maternally inherited, which means that the bacteria are found in the eggs of their hosts, although there is some evidence suggesting that the bacteria can be transmitted between hosts as well (Baldo et al., 2008). Wolbachia have a mutualistic relationship with their nematode hosts, whereas they act as reproductive parasites in arthropods. Wolbachia are extraordinarily prevalent in insects; it is estimated that over 65% of all insects harbor Wolbachia (Hilgenboecker et al., 2008).

Recently, Wolbachia has received a lot of attention for its medical applications. Wolbachia are being used as drug targets for treatment of lymphatic filariasis, a neglected tropical disease (Casiraghi et al., 2002). Wolbachia are mutualists of the filarial nematodes responsible for lymphatic filariasis, where they are necessary for host reproduction. As a result, they are possible drug targets for the therapy for the disease. Indeed, the use of antibiotics to deplete the worms of their Wolbachia has been demonstrated to be an effective chemotherapeutic measure for lymphatic filariasis. Wolbachia also have applications in insect vector control (Hoffmann et al., 2011). Experiments are underway to
use *Wolbachia* to control mosquito populations to limit the spread of dengue fever. In an evolutionary context, perhaps *Wolbachia*’s medical relevance is not so surprising given that *Wolbachia* are also relatives of the Rickettsial human pathogens *Anaplasma* and *Ehrlichia* (Darby et al., 2007)

*Wolbachia* are best known for the range of reproductive abnormalities they cause in their insect hosts. *Wolbachia* have been documented to induce male-killing, parthenogenesis induction, feminization of male offspring (feminization), and sperm-egg incompatibility between infected and uninfected insects, also known as cytoplasmic incompatibility (CI). In CI, infected females can produce viable offspring with both infected and uninfected males, whereas uninfected females can only produce viable offspring with uninfected males (reviewed in Werren, Baldo et al, 2008). Currently, the exact molecular mechanisms behind these reproductive manipulations are unknown. All four of these reproductive manipulations (male-killing, parthenogenesis, feminization, and cytoplasmic incompatibility) serve to increase the fecundity of infected females in the population. As *Wolbachia* are transmitted maternally, this means increased numbers of *Wolbachia*.

**Cytoplasmic Incompatibility**

Sperm-egg incompatibility, or cytoplasmic incompatibility (CI) is the most common of the reproductive manipulations caused by Wolbachia. In CI, crosses between uninfected females and infected males result in abnormal embryos. In CI embryos, the male pronucleus lags behind the female pronucleus during the first mitotic division. This anomaly is readily apparent during metaphase, where the paternal chromatin is not fully
condensed and the maternal chromosomes are. During anaphase, separation results in either fragmentation or complete exclusion of the male pronucleus. Often times, a chromatin bridge can be observed connecting the two daughter cells during telophase (Ryan and Saul, 1968). Infected females, however, can mate with both uninfected and infected males and thus produce more offspring than uninfected females. As infected females can pass *Wolbachia* onto their progeny, CI serves to propagate Wolbachia throughout the host population.

The molecular mechanisms of CI are not known. However, most researchers agree that *Wolbachia* somehow modify the sperm. This modification can be rescued if there are *Wolbachia* present in the egg (Werren, 1997). Currently, there are three mathematical theoretical predictions based on experimental data: 1) the lock-key hypothesis 2) the titration-restitution hypothesis, and 3) the timing hypothesis. The lock-key hypothesis postulates that *Wolbachia* produces a factor that “locks” the male pronucleus during spermatogenesis (Breeuwer and Werren, 1990; Hurst, 1991; Poinsot and Mercot, 1999; Werren, 1997). This “lock” can only be removed by a “key” made by *Wolbachia* in the egg. Thus, according to the lock-key hypothesis, uninfected females are unable to mate with infected males as they do not have *Wolbachia* to produce the “key” to unlock the male pronucleus. Importantly, the lock and key hypothesis also explains incompatibilities between different *Wolbachia* strains (often referred to as bidirectional incompatibility), as infected females cannot mate with males infected with a different *Wolbachia* strain. This intuitively makes sense as the “key” produced by one strain of *Wolbachia* would not be able to unlock a “lock” made by another (Poinsot et al., 2003). Currently, there are no proposed hypotheses as to which protein factors act as the “lock” or the “key.”
The titration-restitution hypothesis postulates that a factor is removed from the male pronucleus during spermatogenesis (Kose and Karr, 1995). This factor is later restored by Wolbachia in the egg upon fertilization. Thus, according to the titration-restitution hypothesis, uninfected females cannot mate with infected males as they do not have Wolbachia to restore the missing factor. The titration-restitution hypothesis can also explain bidirectional incompatibilities. According to the titration-restitution hypothesis, each Wolbachia strain can either 1) remove a different DNA-binding host protein or 2) has its own different titration-restitution profile. It has been suggested that Hira, a histone H3.3 chaperone, is the protein factor in question (Zheng, 2011 #25). It is unclear which of the two hypotheses is correct.

The timing hypothesis postulates that CI is caused by Wolbachia-induced delays of the host cell cycle, specifically in the male pronucleus. Unlike the previous two hypotheses reviewed, which are supported by ecological experimental crosses between infected individuals, the timing hypothesis is corroborated by cell biological experimental evidence (Tram and Sullivan, 2002). Most studies of CI have been done on Drosophila simulans or Nasonia vitripennis. In uninfected embryos, upon fertilization, the paternal pronucleus (the paternal contribution of genetic material) undergoes modifications dependent on the maternal pronucleus. Paternal DNA is modified in such a way that proteins required for chromatin condensation (sperm-derived histones) are replaced with maternal counterparts, and a maternally derived nuclear envelope surrounds the newly formed male pronucleus. Meanwhile, the egg is released from metaphase II arrest and completes meiosis. The male and female pronuclei then prepare for entry into mitosis and begin to migrate towards each other (Tram et al., 2003)
However, *Wolbachia* have been shown to cause a delay in nuclear envelope breakdown (NEBD) of the male pronucleus, affecting the ability of the female and male pronuclei to enter into mitosis simultaneously (Tram and Sullivan, 2002). *Wolbachia* may delay the cell cycle by directly interacting with CDK1, an enzyme required for cell cycle entry into mitosis. The mechanism behind this delay was revealed by Tram et al (2002), who showed that *Wolbachia* causes a delay in the phosphorylation of histone H3 in the male pronucleus of CI embryos, a phenotype that is correlated with CDK1 inactivation, suggesting that *Wolbachia* may produce a protein factor that directly interferes with cell cycle checkpoints. According to the timing hypothesis, infected females can produce viable offspring with both uninfected and infected males because the female pronucleus is correspondingly delayed in these embryos, restoring male and female pronuclei synchrony (Tram and Sullivan, 2002).

In CI embryos, defects in the male pronucleus are apparent at interphase, the cell cycle phase during which the cell prepares for entry into mitosis. During interphase, paternal protamines (proteins that allow for more compact chromosome condensation in sperm) are substituted with maternal H3.3 histones, and the nucleosome complex forms. In CI embryos, a delay in H3.3 and H4 deposition is observed and although S phase is initiated at the same time in the male pronucleus as in the female pronucleus, DNA replication is delayed in the male pronucleus in CI embryos, continuing into metaphase (Landmann et al., 2009). One big problem with the timing hypothesis is that although it explains the reason why uninfected females cannot mate with infected males, it does not explain the incompatibilities between gametes coming from hosts that are infected with different *Wolbachia* (although it is possible that different strains delay the cell cycle to
different extents), nor can it explain how infected females can mate with uninfected males, if there is still an asynchrony between maternal and paternal pronuclei (Poinsot et al., 2003).

Other research suggests that Wolbachia-induced cell cycle delays are a consequence, and not the cause, of CI. For example, Wolbachia inhibits androgenic development in Drosophila CI embryos (Ferree and Sullivan, 2006). In androgenetic development, zygotes form from the fusion of two paternal pronuclei as opposed to one paternal pronucleus and one maternal pronucleus (Komma and Endow, 1995). If the timing hypothesis were correct, androgenic CI embryos should develop normally as there should be no asynchrony between the two pronuclei. If, however, the lock-key or titration-restitution hypothesis were correct, then androgenic development would be inhibited because of implied Wolbachia-induced defects in the male pronucleus. Indeed, the frequency of androgenic development is significantly lower when sperm come from infected males than in normal controls, suggesting that CI is caused by an inherent Wolbachia-induced defect in the paternal pronucleus rather than cell-cycle asynchrony between the male and female pronuclei. These defects are attributed to Wolbachia-induced modifications during spermatogenesis (Ferree and Sullivan, 2006). Thus, the lock-key hypothesis and the titration-restitution hypothesis are better models for CI than the timing hypothesis. As experimental evidence from different Wolbachia-host models is contradicting, it is clear that more research needs to be done before the exact molecular mechanisms of CI can be determined.

**Type IV Secretion Systems**
How is *Wolbachia* manipulating the host to cause these reproductive effects? One possible avenue is through secretion of factors into the host. In one study, infected *N. vitripennis* females were cured of their *Wolbachia* infection with the antibiotic tetracycline. Although these females no longer had their *Wolbachia*, they were still able to produce viable offspring when mated with infected males, suggesting that the bacteria, although no longer present, may have secreted a factor that can continue to rescue CI in their absence (Breeuwer and Werren, 1993). More evidence for proteins secreted by *Wolbachia* comes from the modification of sperm; although *Wolbachia* are not found in the developing sperm, they are capable of modifying the sperm of the beetle *Chelymorpha alternans*, which suggests that the bacterium secretes extracellular factors to manipulate the biology of its hosts (Clark et al., 2008). Similar observations have been made in the mutualist *wBm*, the *Wolbachia* of *B. malayi*, where the effects of *Wolbachia* depletion have been observed (Albertson et al., 2009). Upon clearing the *Wolbachia* from the filarial nematodes, little to no apoptosis was observed in the cells closest to the *Wolbachia* infection (adjacent lateral hypodermal chord cells). However, extensive apoptosis was observed in the adjacent female germline cells, embryos, microfilariae and L4 larvae, which contain very few *Wolbachia*. These results led the authors to speculate that *Wolbachia* may secrete a factor that prevents apoptosis in these cells (Albertson et al., 2009), again pointing to the possibility of a *Wolbachia* secretion system.

Indeed, this experimental evidence is supported by the sequencing of the *Wolbachia* genome, which was found to encode a Type IV secretion system (Wu et al., 2004). Type IV Secretion Systems (T4SSs) are large, multi-protein apparati that span the bacterial membrane (Fronzes et al., 2009). T4SSs secrete proteins (among other substrates, like
DNA-protein complexes), called effectors, into target cells (other bacteria or eukaryotes, including plants and humans) (Backert and Meyer, 2006; Ding et al., 2003; Heidtman et al., 2009; Kagan and Roy, 2002; Luo and Isberg, 2004). These secreted proteins are called “effectors” for the effect they have on their target cells (Ding et al., 2003) and are well known for their important roles in bacterial pathogenesis. For example, the bacterium *Helicobacter pylori*, which has been linked to the development of gastric carcinoma in humans, uses a T4SS to translocate the effector protein CagA into gastric epithelial cells. CagA is important in *H. pylori* pathogenesis; it aberrantly activates β-catenin, which in turn upregulates genes involved in carcinogenesis (Franco et al., 2005). *Legionella pneumophila*, the causative agent of Legionnaire’s Disease, also uses T4SSs and effectors to modify the biology of its human host. For example, *Legionella* uses effector proteins to recruit endoplasmic reticulum secretory vesicles (Kagan and Roy, 2002).

As they have evolved in an eukaryotic context, to interact with eukaryotic proteins, effectors tend to have distinctive characteristics that are easily identified using bioinformatics. For example, effectors 1) have eukaryotic homology or 2) contain eukaryotic domains (Lesser and Miller, 2001).

**The Wolbachia Genome**

In order to understand the *Wolbachia* genome, we must first understand bacterial genomics in general. Bacterial genomes are located in a circular chromosome in the cell and are much smaller than eukaryotic genomes. Unlike eukaryotic genomes, bacterial chromosomes are compact, containing very little non-protein-coding DNA.
Bacterial endosymbionts, like *Wolbachia*, tend to have smaller genomes than other bacteria. The general trend for obligately intracellular endosymbionts is genome reduction; these bacteria tend to lose genes and they do not have any means of obtaining new genes through horizontal gene transfer from the environment.

In keeping with this reductive trend, the *Wolbachia* genome is 1.08-1.7 Mb. Unlike most endosymbiont genomes, however, the *Wolbachia* genome has undergone many genomic rearrangements, inversions, and deletions (Wu et al., 2004), most likely because the bacteria form parasitic relationships with their arthropod hosts and must continually evolve to adapt to the hostile host environment (which selects against them). It also contains many mobile DNA elements, including an active phage (Newton and Bordenstein, 2011). The genome also contains many repetitive elements, most of which encode for hypothetical proteins or ankyrin domains, a primarily eukaryotic domain that is involved in protein-protein interaction (Ishmael et al., 2009).

**The Power of Bioinformatics**

Currently, there is no tractable genetic system for *Wolbachia*, making bioinformatics an attractive tool for studying *Wolbachia* biology. Bioinformatics is the application of computer science to biology and, as an umbrella term, encompasses many aspects of biology including genomics (of relevance to this work). When a genome is sequenced, it is structurally annotated (finding coding regions) and then functionally annotated (predicting functions for the encoded proteins). For example, we can use start and stop codons to predict open reading frames (genes) in a DNA sequence (using the
program getORF, for example). This information can then easily be translated into an amino acid sequence, using a codon mapping matrix, opening up avenues for further study. For example, one could look at the chemical properties of the amino acids to determine the relative hydrophilicity or hydrophobicity of the protein. In addition to predicting protein function, one could also tease out the evolutionary relationships between the two proteins.

As mentioned above, bioinformatics can also be used to predict a protein’s function – the functional annotation of a genome. The function of a protein is related to its structure and homology. Proteins with similar structures tend to have similar functions (catalytic, enzymatic, or otherwise) and are called homologs. Thus, one could predict a particular protein’s function by looking for homology to other proteins or protein domains. This is particularly useful for predicting the functions of unknown, or newly discovered, proteins. Indeed, there exist many gene and protein databases, such as the National Center for Bioinformatics (NCBI)’s Genbank and the Wellcome Sanger Trust’s Pfam (Protein family) databases. Both of these, utilized herein, are searchable with well-established algorithms such as Basic Local Alignment Search Tool (better known by its acronym, BLAST) and HMMer (Hidden Markov Modeler). Searching these two databases with genes from Wolbachia genomes will tell us something about the homology of these proteins to other known proteins and what organism the closest known homolog comes from.

A Yeast Genetic Screen for Bacterial Effectors

The lack of a tractable genetic system has made studying Wolbachia pathogenesis extremely difficult. However, more than 10 years ago, Dr. Cammie Lesser (2001)
developed an assay that makes it possible to study effectors of obligately intracellular bacteria, such as *Wolbachia*, using the basal eukaryote *Saccharomyces cerevisiae* (budding yeast) as a model for host infection. *S. cerevisiae* is an excellent system in which to study bacterial pathogenesis as most bacterial effectors target core eukaryotic processes that are conserved in yeast and higher eukaryotes (Lesser and Miller, 2001; Siggers and Lesser, 2008; Slagowski et al., 2008). These cellular processes are heavily studied in *S. cerevisiae*, making yeast well-suited as a model for bacterial effectors. What is more, yeast are easy to grow and maintain in the laboratory. They are also genetically tractable. Over 75% of yeast open reading frames (putative genes) have known or predicted functions. There are an extraordinary number of strain banks and genetic tools that have been developed to facilitate the study of their biology. Most of these tools can also be applied in the Lesser and Miller yeast genetic screen, increasing its versatility.

In the assay, candidate effectors are ectopically expressed inside *S. cerevisiae* cells, which is exactly where they would have been had they been injected by the bacteria into the cell *in vivo*. As effectors are designed to interact with eukaryotic proteins and processes, their expression interferes with normal cell processes and causes a measurable growth defect in yeast that can be quantified by growing the yeast in liquid media and measuring the turbidity of the solution (Kramer et al., 2007; Lesser and Miller, 2001; Siggers and Lesser, 2008). Only effector proteins cause growth defects in yeast; other bacterial proteins do not cause growth defects in yeast (Arnoldo et al., 2008; Campodonico et al., 2005; Kramer et al., 2007).

In addition to identifying bacterial effectors, the Lesser and Miller (2001) yeast genetic screen can also determine effector localization in the eukaryotic cell. In the assay,
candidate effectors are fused to green fluorescent protein (GFP). This makes it possible to determine their localization in the eukaryotic cell via fluorescent microscopy. Many studies have demonstrated that effectors localize to the same parts of the yeast cell as they do in their hosts (Lesser and Miller, 2001; Slagowski et al., 2008).

The assay is not perfect, however. To begin, the screen cannot detect effectors that target host-specific processes that are not conserved in yeast. In addition, effectors are not expressed in the same levels as they are in vivo.

**Overview of Study**

In this study, we sought to identify and characterize the Type IV effector proteins *Wolbachia* use to manipulate the reproduction of its insect hosts. We specifically looked at *wVitA*, the *Wolbachia* strain that infects the parasitic wasp, *Nasonia vitripennis*, which was used to study cytoplasmic incompatibility. We first searched predicted *wVitA* genes against the databases Genbank and Pfam (using BLAST and HMMer) to look for conserved eukaryotic domains and homology. Genes encoding proteins with significant eukaryotic homology were considered candidate effectors. We also included in our set of candidate secreted effectors genes of unknown function that are specific to *Wolbachia*, that is, not found in any other genome, for species-specific genes are often effectors (Ralph Isberg, personal communication). These candidate effectors were then screened for growth defects in yeast using the Lesser and Miller assay, and their localization in the yeast cell was characterized in the hopes of better understanding their biochemical function.
Methods

**Identification of candidate wVitA effectors using bioinformatics.** Open reading frames (ORFs) for the wVitA genome were determined by getorf (of the EMBOSS program suite). Getorf uses a simple model for gene calling based on start and stop codons (http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html). Each ORF was used as a query to both Genbank’s nr database (using Blastp) and PFam (using HMMer). Blastp looks for sequence homology between each predicted ORF and the public sequence archive while HMMer uses hidden markov models to find conserved domains within the predicted ORFs and annotated in PFam. The output from these bioinformatics analyses was parsed with in-house Perl pipelines to filter for eukaryotic hits using a taxonomy database generated by NCBI and by PFam. Hits were sorted based on significance, with the most significant (E=0.0001) comprising the pool of candidate effectors.

**Amplification of candidate wVitA effectors.** Candidate wVitA effectors were amplified via PCR using Phusion® High-Fidelity PCR Master Mix with HF Buffer (Thermo Scientific). Success of the amplification as well as amplicon size was verified using gel electrophoresis to ensure that the correct ORF had been amplified. Verified PCR products were purified using the QIAquick PCR purification kit (Qiagen) per manufacturer’s instructions.

**Insertion into pENTR™ entry vector.** The DNA concentration of the purified PCR products was first verified using a NanoDrop1000 before ligation into pENTR™ vectors (Methods Figure 1) using the pENTR/D-TOPO® cloning kit (Invitrogen). Resultant cloning reactions were used in the transformation of One Shot® TOP10 Chemically Competent *Escherichia coli* cells (Invitrogen). As the pENTR plasmid encodes for kanamycin
resistance, the transformants were plated onto selective LB/KAN media (0.02 mg/ml kanamycin) to ensure that only cells that took up the plasmid would grow. Four representative transformants for each of 32 transformation reactions were selected to be sequenced via the Sanger method using the M13F sequencing primer at Beckman Coulter Genomics. After resultant sequences were obtained, they were blasted against the wVitA assembly to ensure the identity of the ORF. In addition, sequences were screened by eye to make sure that the att cloning sites were in frame with the candidate effector.

**Methods Figure 1. The pENTR entry vector.** Figure generated by Invitrogen.

**Insertion into pFus destination vector (yeast expression vector).** pENTR plasmids that had been determined to contain candidate effectors in the correct reading
frame were extracted from *E. coli* cells using the QIAprep 96 Turbo Miniprep Kit (Qiagen) per manufacturer’s instructions. The ORFs in these plasmids were used in the LR (Methods Figure 2) reaction with a pFus destination plasmid (Methods Figure 3) using the Gateway\textsuperscript{TM} LR Clonase\textsuperscript{TM} II Enzyme Mix (Invitrogen) per manufacturer’s instructions. The LR reactions were then transformed into DH5-α *E. coli* cells. As the pFus expression plasmid encodes for ampicillin resistance, the transformants were plated onto LB/AMP media (0.1 mg/ml ampicillin) to ensure that only those cells that took up the plasmid would grow. Again, four representative transformants for each of 32 transformation reactions were selected for Sanger sequencing using a sequencing primer at Beckman Coulter Genomics. After resultant sequences were obtained, they were blasted against the wVitA assembly to ensure the identity of the ORF and screened by eye to make sure that the N terminal GFP tag was in frame with the candidate effector. The pFus galactose-inducible expression vector is a 2 μ plasmid with a leucine metabolic marker and an ampicillin resistance cassette.

![Methods Figure 2. The LR reaction.](image)

Figure generated by Invitrogen.
Methods  

Figure 3. The pFus destination plasmid. Figure provided by Dr. Irene Newton.

Transformation of yeast. Verified plasmids were extracted from the *E. coli* cells per manufacturer's instructions using the QIAprep 96 Turbo Miniprep Kit (Qiagen). The extracted plasmids were then transformed into wildtype yeast (strain BY4741), and the transformants were plated onto selective yeast media lacking leucine (YM-leu 2% D).

Screen for effectors. In order to induce expression of our candidate effectors, transformed yeast were grown in YM-leu 4% galactose liquid and solid media at 30°C to
screen for growth defects. As a control, growth of the same strains in 2% dextrose (glucose) media was also monitored as was growth of the yeast harboring vector alone (pFus). For monitoring growth in liquid media, OD$_{600}$ measurements were taken at 0 hr, 24 hr, and 72 hr post induction. A serial dilution assay (using 1:10 dilutions) was performed to confirm all growth defects seen in the liquid assay. In this assay, yeast that showed a growth defect in liquid were plated onto solid 2% dextrose and 4% galactose YM-leu plates and left to incubate for 48 hours at 30°C.

**Localization in the eukaryotic cell.** Transformants expressing candidate effectors were allowed to grow in the presence of galactose for 48 hours before they were fixed in 70% ethanol and visualized on a Nikon 800 scope at 100x objectives.
Results

A total of 295 candidate wVitA effectors were identified using bioinformatic methods.

We used bioinformatics methods to identify possible secreted wVitA effectors. First, a newly sequenced wVitA draft genome assembly was structurally annotated; open reading frames were identified with the program getORF (http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html). In order to identify open reading frames (ORFs) with eukaryotic characteristics, we conducted a BLASTp and an HMMer search of both Genbank and PFam to pick out wVitA ORFs that 1) had eukaryotic homology (meaning a high score in blast against a sequence from a eukaryote) or 2) contained eukaryotic domains (a significant PFam score against a domain normally found only in eukaryotes). Using an in-house Perl pipeline (provided by Dr. Irene Newton), we were able to find a total of 295 ORFs that fit these criteria. A total of 138 ORFS were identified using HMMer, and 181 ORFs were identified using Blast; 24 of these ORFs were found in both subsets of data. As these ORFs had eukaryotic homology or contained eukaryotic domains, we termed them Wolbachia Eukaryotic-like Genes or WEGs.

As these WEGs are hypothesized to be the result of a horizontal gene transfer into the Wolbachia genome we were curious to explore the taxonomic distribution of eukaryotic phyla to which these candidate effectors had homology. A broad range of eukaryotic phyla was represented, ranging in complexity from unicellular organisms, like protists, to multicellular ones, like the chordates (Results Figure 1). The distribution of eukaryotic
phyla for the 181 BLAST WEGs indicates that no one phylum constituted a majority of the hits (Results Figure 1). The single most abundant phylum, fungi, only accounted for 23% of the hits. Surprisingly, *Wolbachia*’s hosts, the phyla *Arthropoda* and *Nematodea* did not account for a majority of the hits, accounting for only 8% and 2% of the hits, respectively.

**Results Figure 1. Phylum-level taxonomic homology between wVitA candidate effectors and Genbank reveals a diversity of eukaryotic homologies.** Blast search output for each wVitA ORF was filtered for hits to eukaryotes using the ncbi taxonomy. Eukaryotic hits were grouped by phylum and assembled into a pie chart using Microsoft Excel 2008. No one phylum accounted for a majority of the hits. *Wolbachia*’s host phyla, *Arthropoda* and *Nematodea*, only account for 8% and 2% of the hits, respectively.

**Comparative genomic analysis.**

*Wolbachia* strains are quite numerous and divergent with regard to the phenotypes they induce in their host. Therefore, in order to study any possible effects between host-induced reproductive effect and WEG content, a comparative genomic analysis was
conducted (Figure 2). The genomes of wVitA, wUni, wPip, wMel, and wBm were compared with regard to their candidate effector content, based on PFam search output (which gives a sense of the eukaryotic domains within each genome). All five genomes were subjected to the same bioinformatics pipeline described above.

A total of 51 domains were found in common in at least two of the five genomes queried (Results Figure 2). Out of these 51 domains, only 12 were found across all five genomes: Ank, bac surface Ag, CLP protease, HHH, KH_1, KH_2, Maf, SRP SPB, SRP54, SRP54 N, Tim44, Trypsin, and YchF-GTPaseC. The wBm genome encoded the fewest total number of domains. Of relevance to this analysis is the mutualistic relationship between wBm and its nematodee host.

The wPip, wMel, and wVitA parasites all induce cytoplasmic incompatibility (CI) so we were curious to investigate the shared set of effector domains within this group. There was no correlation between domain prevalence and reproductive phenotype. However, there was one domain that was found in all three genomes. The single domain found across all CI inducing strains but not in any other Wolbachia in this analysis was Sua5 yciO yrdC, suggesting that this particular domain may be relevant for the CI phenotype. Sua5 yciO yrdC is a dsRNA binding domain {Teplova, 2000 #26}. It has been implicated in tRNA threonylcarbamoyladenosine biosynthesis pathway in both eukaryotes and prokaryotes {Yacoubi, 2009 #27} and has also been found to play a role in telomere recombination in yeast {Lin, 2010 #28}. 
Investigating the biochemical functions of wVitA type IV effectors

Park 24

 Ank
 Anoctamin
 Apolipoprotein
 ATG16
 Bac surface Ag
 Bir1p
 Clp N
 CLP protease
 Collagen
 COMPASS-Shq1
 DEAD
 DBP81
 dem
 Dynein heavy
 ERM
 Guanylate kin
 HHH
 KAP NTPase
 KH 1
 KH 2
 Mecolin
 Maf
 Mfa12 component
 MMR HSRT
 Mod r
 Myosin tail 1
 NB-ARC
 NDK
 OTU
 PDA0
 S1
 S4
 Sdh5
 SF-assemble
 SMC N
 SNF2 N
 Spt4
 SRP SPB
 SRP54
 SRP54 N
 Suu5 yO yzC
 Surface Ag 2
 TBP5P
 TEBP beta
 TIm44
 TLC
 TMF DNA bd
 Trypsin
 Tubulin
 WH2
 YehF-GTPase C
Results Figure 2. The number of eukaryotic domains found (through a Pfam search) in a diverse group of Wolbachia genomes as represented by a bubble chart. Candidate effectors from each of five sequenced genomes were used as queries to Pfam and significant hits to eukaryotic domains wUni (in red), wPip (in orange), wMel (in yellow), wBm (in green), and wVitA (in blue) are shown. The size of each bubble represents relative number in each category. Reproductive effects induced do not correlate with eukaryotic domain content as wPip, wMel, and wVitA all display CI but clearly differ with regards to genomic profile as highlighted here. The 5 species share 12 domains in common highlighted in bold lettering.

Candidate wVitA effectors caused growth defects in yeast.

We used an assay developed by Lesser and Miller (2001) to identify wVitA effectors using yeast as a model for eukaryotic infection. True bacterial effectors have evolved within the eukaryotic cell – to interact with eukaryotic cell processes; they cause a measurable growth defect in yeast, compared to bacterial core proteins. Using this well-established assay, we screened a total of 34 wVitA candidate effectors. Wild type yeast (strain BY4741) previously transformed with candidate wVitA effectors encoded in a galactose inducible plasmid were grown in the presence (4% galactose) or absence (2% glucose) of induction substrate. To measure if expression of the wVitA candidate effector had a measurable effect on growth, the optical density (OD$_{600}$) of the resulting cultures was measured 0, 24, 48, and 72 hours post inoculation. Demonstrative growth defects were observed for 10 out of 34 of these candidate effectors after 24 hrs of expression (Results Figure 3). These growth defects observed in liquid culture were recapitulated on solid media; 1:10 serial dilutions performed on solid YM-leu agar confirmed the observed growth defects based on optical density (Results Figure 4). WEGs A4, A6, B2, B6, B8, C3, C4, C5, C9, and C10 caused growth defects in yeast. The vast majority of WEGs (including ones that did not cause growth defects) had a cytoplasmic with punctate localization (Results Table 1). Surface antigen did not cause a growth defect in yeast. Several ankyrin-domain containing proteins
were screened in this study including C5, WEG 00037_5 (data not shown), and WEG 00090_3 (data not shown). Out of these three, only C5 caused a measurable growth defect in yeast. Of particular interest is the domain KH_1, a eukaryotic nucleic acid binding domain. One WEG containing this domain, B2, caused a growth defect in wildtype yeast. The Sua5_yciO_yrdC domain was unique to the three Wolbachia species that cause cytoplasmic incompatibility (Results Figure 2). One Sua5_yciO_yrdC containing protein (WEG 00048_13, data now shown) was screened and did not cause a growth defect in wildtype yeast. Finally, two ribosome-binding GTPase domain-containing proteins were included in this screen. Interestingly, B8, which encodes for a protein that has homology to a 50S ribosome-binding GTPase, caused a growth defect in wildtype yeast. However, a similar protein, WEG 00013_16 (data not shown), which also encodes for a 50S ribosome-binding GTPase, did not cause a growth defect in wildtype yeast.
Figure 3. Growth defects in yeast expressing wVitA candidate effectors as observed by optical density measurement. Yeast were grown in 2% glucose and 4% galactose liquid YM-leu media at 30°C. The optical density (OD) of the yeast was measured at 24 hours and 10/34 candidates exhibited significant growth defects compared to an average of the ODs of all yeast, including those that did not cause growth defects.
Investigating the biochemical functions of wVitA type IV effectors

Figure 4. Growth defects in yeast expressing wVitA candidate effectors as observed by spotting assay on solid agar plates. A serial dilution assay was performed to confirm all growth defects seen in the liquid assay. Yeast harboring candidate effectors on a galactose inducible plasmid were grown to confluence and then optical density was normalized to OD$_{600}$ = 1.0 before 1:10 dilutions were spotted onto solid 2% dextrose or 4% galactose YM-leu agar. After a 24 hour incubation at 30°C plates were imaged. Yeast expressing vector alone were used as a control. Top: 2% dextrose, Bottom: 4% galactose

Localization of candidate type IV wVitA effectors in yeast

In addition to allowing us to identify candidate Type IV effectors, the Lesser and Miller (2001) assay also allows us to characterize their localization in the eukaryotic cell. Cloned WEGs were fused to GFP at the N-terminus, which allowed us to use fluorescence microscopy to visualize their localization in the cell. We focused our localization analysis
on the 10/32 WEGs that caused significant growth defects in yeast (Results Figure 2).
Yeast expressing these WEGs were grown for 24 hours in YM-leu 4%G to induce expression before fixation in 4% paraformaldehyde. In addition, and to explore the effects of an alternative fixation on the localization signal, 48 hour cultures of all WEGs (32 total) were grown in 4%G media and fixed in ethanol. Yeast were visualized using a fluorescent microscope.
Results Table 1. Summary of WEGs that caused growth defects in wildtype yeast.

<table>
<thead>
<tr>
<th>WEG</th>
<th>contig</th>
<th>Domain</th>
<th>Localization (48 hr)</th>
<th>Pictures (48 hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A4</td>
<td>00137_9</td>
<td>Peptidase_M50</td>
<td>Cytoplasmic with punctate localization</td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>00009_2</td>
<td>Hypothetical protein in <em>Trypanosoma cruzi</em></td>
<td>Cytoplasmic with punctate localization</td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td>00043_1</td>
<td>KH_1 (eukaryotic nucleic acid binding domain that can bind either RNA or single-stranded RNA)</td>
<td>Cytoplasmic with punctate localization</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>B6</td>
<td>00086_11</td>
<td>DPBB_1 (function not well understood)</td>
<td>Cytoplasmic with punctate localization</td>
<td></td>
</tr>
<tr>
<td>B8</td>
<td>00086_6</td>
<td>MMR_HSR1 (50S ribosome-binding GTPase)</td>
<td>Unknown (no picture taken)</td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>00014_2</td>
<td>Hypothetical protein in <em>Branchiostoma floridiae</em></td>
<td>Cytoplasmic with punctate localization</td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>00017_15</td>
<td>GTP_EFTU (GTP-binding elongation factor)</td>
<td>Nuclear/ER</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Description</td>
<td>Localization</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-------------------------</td>
<td>-------------------------------</td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>00090_7</td>
<td>Ank</td>
<td>Cytoplasmic with punctate localization</td>
<td></td>
</tr>
<tr>
<td>C9</td>
<td>00117_1</td>
<td>Hypothetical protein found in <em>Trichomonas vaginalis</em></td>
<td>Cytoplasmic with punctate localization</td>
<td></td>
</tr>
<tr>
<td>C10</td>
<td>00121_3</td>
<td>lysozyme found in bacterial genus <em>Leifsonia</em></td>
<td>Cytoplasmic with punctate localization</td>
<td></td>
</tr>
</tbody>
</table>


**Discussion**

We set out to identify and characterize the molecular mechanisms the obligately intracellular bacteria *Wolbachia* use to manipulate their eukaryotic arthropod and nematode hosts. As the *Wolbachia* genome encodes for a Type IV secretion system, we hypothesized that *Wolbachia* may be manipulating its eukaryotic hosts by secreting Type IV effectors. Thus, we set out to identify and characterize the functions of *Wolbachia* Type IV effectors. We chose to focus on wVitA, the *Wolbachia* species that infects the parasitic wasp, *Nasonia vitripennis*, wVitA, which has been used to study cytoplasmic incompatibility.

**Bioinformatics**

Using a pipeline provided by Dr. Irene Newton, we used BLASTp and HMMer to identify predicted wVitA ORFs that 1) had significant sequence homology to eukaryotic proteins (using BLASTp to query Genbank) or 2) contained significant sequence homology to eukaryotic domains (using HMMer to query PFam). A total of 295 wVitA ORFs were identified that had significant eukaryotic homology or contained eukaryotic domains. 181 of these ORFs were identified using BLASTp.

Only 10% of the BLASTp identified WEGs have homology to the phyla arthropoda and nematodea seem to target arthropod or nematode specific processes. It is unknown whether these WEGs target processes that are unique to arthropods or nematodes or whether they target processes that are common to all eukaryotic cells. If the former is true, this would stress the importance of the host background in determining the phenotype.
The remaining 138 WEGs were identified using an HMMer driven search of the Pfam database, which was used to identify ORFS that contained eukaryotic domains. We also compared the Pfam results from wVitA to those from wRi (parasitic wasp Muscidifurax uniraptor), wPip (found in the mosquito Culex pipiens), wMel (found in Drosophila melanogaster), and wBm (found in the filarial nematode Brugia malayi), and wVitA (Results Figure 2).

The wBm genome had the smallest total number of domains. Interestingly, wBm is also the only one of the five Wolbachia compared that forms a mutualistic relationship with its host; the other four bacteria species act as reproductive parasites in their hosts. This suggests that the additional WEGs found in the parasitic Wolbachia strains may be responsible for the Wolbachia-induced reproductive parasitisms in arthropods. Unlike their nonpathogenic relatives, pathogenic bacteria contain additional genes required for virulence. These genes are encoded on pathogenicity “islands,” which are acquired via horizontal transfer (Hueck, 1998). It is possible that the genes responsible for Wolbachia-induced reproductive parasitisms, such as cytoplasmic incompatibility, are encoded on pathogenicity islands that are transferred horizontally between different Wolbachia.

A total of 51 eukaryotic domains were found in at least two of the five genomes. A total of 12 domains (ank, bac surface Ag, CLP protease, HHH, KH_1, KH_2, Maf, SRP_SPB, SRP54, SRP54_N, Tim44, Trypsin, and YchF-GTPaseC) were found across all five genomes. The fact that these 12 domains are found in both mutualistic and parasitic Wolbachia strains suggests that they are involved in processes that are common to all Wolbachia, regardless of their relationship to the host.
The ankyrin repeats deserve a special mention. Ankyrin repeats motifs are involved in protein-protein interactions, where they act as a scaffold for further molecular interactions. Ankyrin repeats are found across all three domains of life, although they are far more common among eukaryotes; they have even been found in viruses (Bork, 1993 #29). Ankyrin repeats are found in a diverse array of proteins, ranging from proteins that participate in signaling pathways to ion channels, just to name a few (Mosavi et al., 2004).

The *Wolbachia* genome contains an unusual number of ankyrin repeats for a prokaryotic genome. It has been proposed that these ankyrin repeat containing proteins are involved in mediating *Wolbachia*-host interactions (Wu et al., 2004). Indeed, studies suggest that this may be the case, making ankyrin repeat containing *Wolbachia* proteins promising Type IV effector candidates. Sinkins et al (2005) hypothesized that the ankyrin repeat containing proteins pk1 and pk2 cause bidirectional incompatibility between two identical *Culex Wolbachia* strains Bei and Pel; in bidirectional incompatibility, Bei females cannot produce progeny with Pel males even though the two strains are identical. What more, Sinkins et al (2005) found that expression of these proteins was sex-specific. They found that pk2 expressed only in females in the pka *Wolbachia* grouping (*Wolbachia* strains Mol, Pel, and Sumo) whereas it was expressed in both males and females of the pkb grouping (Sinkins et al., 2005). This sex-specific expression is suggestive of the modification and rescue model proposed to explain cytoplasmic incompatibility (Werren, 1997). Thus, pk1 may be involved in cytoplasmic incompatibility (Sinkins et al., 2005). All of the above seem to suggest that ankyrin repeat containing *Wolbachia* proteins are involved in modulating *Wolbachia*-host interactions. In this study, two ankyrin-repeat containing WEGs were screened using the Lesser and Miller assay (see below).
Investigating the biochemical functions of wVitA type IV effectors

Park 36

There was no correlation between domain prevalence and reproductive phenotype. wPip, wMel, and wVitA all cause cytoplasmic incompatibility; however, each had a distinct WEG profile. For instance, wPip, wMel, and wVitA have different numbers of ankyrin-repeat containing proteins, as can be seen demonstratively in Figure 2. The wPip genome contains 54 (Duron et al., 2007) and the wAna genome encodes for 34 (Fenn and Blaxter, 2006). This suggests that the host background may have more to do with the reproductive effect than the Wolbachia strain itself.

Lesser and Miller Assay

Next, we set out to determine whether these WEGs were indeed wVitA effectors. As there exists no genetic system in which we can study Wolbachia, we used budding yeast as a model for eukaryotic infection. Candidate WEGs fused to a GFP construct were cloned into galactose-inducible plasmids. These plasmids were then transformed into wildtype yeast. As effectors have evolved to interact with eukaryotic processes, they cause a growth defect in yeast, as compared to core bacterial proteins. The yeast were screened for growth defects with (4% galactose YM-leu media) and without (2% dextrose YM-leu media) the presence of the induction substrate. Optimal density measurements (OD600) were taken to quantify growth at time of inoculation, 24 hours, 48 hours, and 72 hours after inoculation.

Out of a total of 32 assayed proteins, 10 caused growth defects in wildtype yeast (Results Figure 3); these growth defects were recapitulated on solid media. Serial dilutions confirmed these growth defects (Results Figure 4). Based on this data we suggest that
these 10 WEGs are putative wVitA Type IV effectors (Results Table 1). As the WEGs were fused to a GFP construct, we were also able to visualize their localization in the eukaryotic cell by fluorescent microscopy. Wildtype yeast transformed with WEGs were grown overnight in liquid 2% dextrose YM-leu media. Then, they were transferred to 4% galactose YM-leu media and allowed to grow for a period of 48 hours. Afterwards, they were fixed in ethanol and visualized through fluorescent microscopy.

WEG B2, which contained a KH_1 domain, caused a growth defect in wildtype yeast. The eukaryotic KH_1 was present across all five genomes, including wBm, suggesting that it is involved in mediating *Wolbachia*-host processes that are common to both mutualistic and parasitic *Wolbachia*. WEG B2 had a joint cytoplasmic and punctate localization; From the localization pattern of the GFP fusion protein, it appears that WEG B2 appears to co-localize with vesicles, although further experimentation needs to be done to confirm this. The KH_1 domain is a eukaryotic nucleic-acid binding domain. It is not immediately clear why a eukaryotic nucleic-binding domain would be associated with vesicles. Further experimentation needs to be done to elucidate WEG B2’s function in the eukaryotic cell.

Interestingly, B8, which contained a 50S ribosome GTPase domain, caused a growth defect in wildtype yeast. The 50S ribosomal subunit is found in prokaryotes. Although rare, bacterial housekeeping proteins can cause growth defects in yeast. For instance, expression of the housekeeping bacterial *Legionella* protein sterol desaturase caused a growth defect in yeast (Campodonico et al., 2005). However, this is extremely rare. One group of researchers expressed approximately 1000 bacterial *Francisella tularensis* proteins in yeast and found that only 3 caused growth defects (Slagowski et al., 2008).
However, it is also possible that B8 is not a housekeeping protein, but rather has some unknown function in the eukaryotic cell. A similar ribosomal protein, WEG 00013_16, which also encoded for a 50S ribosomal GTPase, did not cause a growth defect (data not shown). Thus, it is possible that B8, although homologous to 50S ribosomal GTPases, is acting in another part of the cell when in the eukaryotic context. However, further experimentation is needed to characterize its biochemical function before before a definitive conclusion can be drawn about WEG B8.

The vast majority of WEGs, including the ones that did not cause growth defects (data not shown), had joint cytoplasmic and punctate localizations. Eight out of 10 growth-defect causing WEGs had a joint cytoplasmic and punctate localization. On first glance, most of the punctate localizations exhibited a pattern typical of actin or vesicular localization. However, co-localization assays need to be done to confirm this. Indeed, bacterial effectors have been shown to affect the eukaryote cytoskeleton (Hueck, 1998). For example, the *Listeria monocytogenes* uses the ActA protein to recruit the eukaryotic Arp2/3 and allow the polymerization of host actin so that it can move around the host cell (Welch et al., 1997). Indeed, one wMel protein appears to be involved in polymerizing actin (Irene Newton, personal communication). It is possible that a similar protein may be found in the wVitA genome.

However, the vast majority of WEGs exhibited a pattern typical of vesicular localization. All except one of the eight growth defect inducing WEGs with punctate localization appeared to localize with vesicles. There are many Type IV bacterial effectors that target vesicles. For example, the *Legionella* effector AnkX is important in inhibiting phagosome-lysosome fusion. AnkX is so named because it contains multiple ankyrin
Investigating the biochemical functions of wVitA type IV effectors

Park

repeats. It has a punctuate localization in the mammalian cell (Pan et al., 2008). As Wolbachia must overcome similar challenges in the host, it is possible that it, too, possesses a similar protein. Indeed, C5, which contains an ankyrin repeat, caused a growth defect in wildtype yeast, making it a putative wVitA Type IV effector. It also had a joint cytoplasmic and punctuate localization. It is possible that it too, like AnkX, could modulate the bacteria’s entry into the host. Ankyrin domains are found across all five analyzed Wolbachia domains, suggesting that some of them must be involved in mediating basic Wolbachia-host processes, such as the inhibition of phagosome-lysosome fusion.

Future Directions.

We hope to finish testing all 295 candidate wVitA effectors in wildtype yeast using the Lesser and Miller assay, as well as perform to actin and vesicle colocalizations assays for the 10 aforementioned WEGs. As GFP tends to aggregate over time, we also hope to get 24 hour images as well to minimize aberrant localizations. In addition, we hope to reperform the assay using yeast genetic mutants. Yeast genetic mutants have been used to identify less phenotypically toxic effectors (Slagowski et al., 2008). Indeed, in a similar screen for wMel Type IV effectors, the vast majority of identified wMel WEGs did not cause a growth defect in the wildtype background. The identified putative actin-polymerizing wMel effector only caused a growth defect in Δswe1 yeast (Irene Newton, personal communication). Thus, it is possible that WEGs that did not cause growth defects in wildtype yeast, such as the two aforementioned ankyrin-repeat containing proteins and the Sua5_ycl0_yrdC domain WEG, are more sensitive and may cause growth defects in mutant
backgrounds. We plan to start with a small subset of yeast deficient in the cell cycle and the protein secretory pathway. In addition, we hope to reperform the growth screen with yeast expressing vector alone.
References


Investigating the biochemical functions of wVitA type IV effectors


