



Published in final edited form as:

Insect Biochem Mol Biol. 2013 September ; 43(9): 867–878. doi:10.1016/j.ibmb.2013.07.002.

Detection of the *Wolbachia* Protein WPIP0282 in Mosquito Spermathecae: Implications for Cytoplasmic Incompatibility

John F. Beckmann and Ann M. Fallon

Department of Entomology; University of Minnesota; St. Paul, Minnesota, 55108, USA.

Abstract

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 *wPip_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 *wPip_0282* and the co-transcribed downstream gene, *wPip_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 *wPip_0282* and *wPip_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.

Keywords

Wolbachia; Cytoplasmic Incompatibility; Mosquito; *Culex pipiens*; 0282; Mass Spectrometry

1. 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 (Jeyaprasak 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

CONTACT: John Beckmann; beckm098@umn.edu; Phone: (507) 412-0807; Fax (612) 625-5299; 1606 Irving Ave N., Minneapolis, MN, 55411. .

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

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 *wPip_0282* in *Culex pipiens* spermathecae as well as in ovarian tissues. In-silico analysis revealed that *wPip_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, *wPip_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.

2. MATERIALS AND METHODS

2.1 Mosquito Colonies and Maintenance

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 *wPip*. 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 *wAlbA* and *wAlbB*) were generously provided by Dr. S. L. Dobson (University of Kentucky).

2.2 Cell lines

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

2.3 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 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.

2.4 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 $\mu\text{m}\times 12\text{ cm}$) packed with C18 resin (5 μm , 200 \AA 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.

2.5 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 $\mu\text{g}/\text{ml}$) 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.

2.6 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.

3. RESULTS

3.1 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 *wPip_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 *wPip_0282* in the annotated genome *wPip* (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.

3.2 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 *wPip* (Pel). The deduced protein sequence [designated *wPip* (Buckeye) in Fig 2] had four amino acid differences over 491 residues (boxed in Fig. 2) relative to the protein encoded by *wPip* (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 *wPip* genomes in *Culex* mosquitoes. The *wPip* 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 *wPip* (JHB)2 in Fig 2] and two partial homologs, one of which [394 residues, ZP_03335681.1, designated *wPip* (JHB)1 in Fig 2] is nearly complete. The third homologous accession (ZP_03335652.1) in *wPip* (JHB) has only 67 residues, and will not be discussed further here. Deduced sequences of *wPip* (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 *wPip* (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 *wPip* (JHB) genome encodes at least two close homologs of *wPip* (Pel)_0282. Moreover, as will be described in further detail below,

BlastP analysis using *wPip* (Pel)_0282 sequence as the subject query uncovered two accessions with ~ 34% identity in *wAlbB* (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, *wPip*_0294 (YP_001975095) in the annotated *wPip* (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 *wPip* (Pel)_0294 homolog sequence as the BlastP subject query, we found that *wAlbB* protein (ZP_09542120.1) had 97% identity, but among JHB accessions, *wPip* (Pel)_0294 had no match closer than *wPip*_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: *wPip* (Pel), *wPip* (JHB), and *wAlbB*, 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 *wPip* (Pel) and shares a pattern of consensus amino acids suggestive of duplication and divergence from an ancestral gene.

3.3 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 *wPip*, *wRi*, *wMel*, *wSim*, *wVitB*, *wHa*, *wNo*, and *wAlbB*; WPIP0282 homologs were consistently absent from the *Wolbachia* genomes of non-CI-inducing, obligate mutualistic nematode strains *wBm* and *wOo*.

Among *Wolbachia* that infect insects, genomes have been fully annotated for *wPip* (Pel), and from two strains that infect *Drosophila*: *wMel* (*Drosophila melanogaster*) and *wRi* (Riverside strain of *Drosophila simulans*). As summarized in Fig 3, the *wRi* genome encodes three well-conserved homologs of *wPip*_0282 while the *wMel* genome encodes only one, *wMel*_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 *wMel* or *wRi* are identical in the two *Drosophila* strains, with the exception of greater divergence in *wRi*_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* 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 *wPip*_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*.

3.4 *w*Pip_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 *w*AlbB (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). 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.

3.5 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 *wPip* (Pel)_0294/0295, were also polycistronic, and transcribed separately from flanking, *wPip* (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).

3.6 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, *wPip* has not been cultivated in cell culture, whereas *wAlbB* has. As shown in Fig 3, *wAlbB* 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 *wAlbB* homologs and performed RT-PCR analysis in both TW-280 cells infected with *wAlbB* and a colony of *Aedes albopictus* mosquitoes dually infected with *wAlbA* and *wAlbB*. Both the 0282 and 0294 homologs were expressed in both cell culture and within the male and female mosquitoes, suggesting that the *wAlbB* 0282 and 0294 homologs are not expressed in a gonadal specific manner (data not shown).

3.7 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 *wPip* (Pel) and their homologs in *wRi* (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 *wPip* 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 *wMel* and *wRi*, we note similar associations with genes that encode phage proteins and ankyrin repeats. Curiously, in *wMel*, there are fewer MGE's in close proximity to the single copy operon, but in *wPip* and *wRi*, 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 *wPip_0283*, are the pseudogenes *wRi_p005380* (nucleotides 573202 to 576723 in NC_012416.1) and *wRi_p010040* (nucleotides 1079661 to 1083182; Fig 7B, indicated by broken lines in the *wRi* 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 *wRi_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.

4. DISCUSSION

4.1 *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 *wPip_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 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

wRi than in wMel, and that wRi causes a more severe CI phenotype than does wMel 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).

4.2 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, wPip_0280, and an integrase gene, wPip_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 wPip_0280 and wPip_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*.”

4.3 Insights from Duplicate Operons and Extrapolations to CI Functionality

We demonstrated by RT-PCR that wPip_0282 is transcribed with wPip_0283 as a polycistronic mRNA, as are the more divergent wPip_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 transferred 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 *wPip* portions of 0283 have been duplicated in the genes *wPip_1291*, and *1292*; the former of which contains an identical copy of the Ulp1 C48 SUMO protease domain harbored in *wPip_0283*. In *wRi*, the two close homologs of this gene are labeled as pseudogenes (*wRi_p005380* and *wRi_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 *wPip_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 *wPip_0282*, that left *wPip_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, *wPip_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 aids 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 toxin-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 toxin 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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 Baldrige, 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.

6. REFERENCES

- Bazzocchi C, Comazzi S, Santoni R, Bandi C, Genchi C, Mortarino M. *Wolbachia* surface protein (WSP) inhibits apoptosis in human neutrophils. *Parasite Immunol.* 2007; 29:73–79. [PubMed: 17241395]
- Beckmann JF, Fallon AM. Decapitation improves detection of *Wolbachia pipiens* (*Rickettsiales: Anaplasmataceae*) in *Culex pipiens* (Diptera: Culicidae) mosquitoes by the polymerase chain reaction. *J. Med. Entomol.* 2012; 49:1103–1108. [PubMed: 23025192]
- Beckmann JF, Markowski TW, Witthuhn BA, Fallon AM. Detection of the *Wolbachia*-encoded DNA binding protein, HU beta, in mosquito gonads. *Insect Biochem. Mol. Biol.* 2013; 43:272–279. <http://www.ncbi.nlm.nih.gov/pubmed/23287400>. [PubMed: 23287400]
- Benjamin B, Yves B, Corinne A. Assembly of the Tc1 and *mariner* transposition initiation complexes depends on the origins of their transposase DNA binding domains. *Genetica.* 2007; 130:105–120. [PubMed: 16912840]
- Bordenstein SR, Reznikoff WS. Mobile DNA in obligate intracellular bacteria. *Nature Rev. Microbiol.* 2005; 3:688–699. [PubMed: 16138097]
- Bourtzis K, Nirgianaki A, Markakis G, Savakis C. *Wolbachia* infection and cytoplasmic incompatibility in *Drosophila* species. *Genetics.* 1996; 144:1063–1073. [PubMed: 8913750]
- Brattig NW, Bazzochi C, Kirschning CJ, Reiling N, Buttner DW, Cecilian F, Geisinger F, Hochrein H, Ernst M, Wagner H, Bandi C, Hoerauf A. The major surface protein of *Wolbachia* endosymbionts in filarial nematodes elicits immune responses through TLR2 and TLR4. *J Immunol.* 2004; 173:437–445. [PubMed: 15210803]

- Callaini G, Dallai R, Riparbelli MG. *Wolbachia*-induced delay of paternal chromatin condensation does not prevent maternal chromosomes from entering anaphase in incompatible crosses of *Drosophila simulans*. *J. Cell Sci.* 1997; 110:271–280. [PubMed: 9044057]
- Clark ME, Veneti Z, Bourtzis K, Karr T. The distribution and proliferation of the intracellular bacteria *Wolbachia* during spermatogenesis in *Drosophila*. *Mech. Dev.* 2002; 111:3–15. [PubMed: 11804774]
- Clark ME, Veneti Z, Bourtzis K, Karr T. *Wolbachia* distribution and cytoplasmic incompatibility during sperm development: the cyst as the basic cellular unit of CI expression. *Mech. Dev.* 2003; 120:185–198. [PubMed: 12559491]
- Cornelis GR, Denecker G. *Yersinia* lead SUMO attack. *Nature Med.* 2001; 7:21–23. [PubMed: 11135606]
- Cubenas-Potts C, Matunis MJ. SUMO: a multifaceted modifier of chromatin structure and function. *Dev. Cell.* 2013; 24:1–12. [PubMed: 23328396]
- Deshaies RJ. SCF and cullin/RING H2-based ubiquitin ligases. *Annu. Rev. Cell Dev. Biol.* 1999; 15:435–467. [PubMed: 10611969]
- Fallon AM, Baldrige GD, Higgins LA, Witthuhn BA. *Wolbachia* from the planthopper *Laodelphax striatellus* establishes a robust, persistent, streptomycin-resistant infection in clonal mosquito cells. *In Vitro Cell. Dev. Biol. Anim.* 2013; 49:66–73. [PubMed: 23271364]
- Fallon AM, Witthuhn BA. Proteasome activity in a naive mosquito cell line infected with *Wolbachia pipientis* wAlbB. *In vitro Cell. Dev. Biol. Anim.* 2009; 45:460–466. [PubMed: 19296184]
- Ferree PM, Frydman HM, Li JM, Cao J, Wieschhaus E, Sullivan W. *Wolbachia* utilizes host microtubules and dynein for anterior localization in the *Drosophila* oocyte. *PLoS Pathog.* 2005; 1:e14. [PubMed: 16228015]
- Frentiu FD, Robinson J, Young PR, McGraw EA, O'Neill SL. *Wolbachia*-mediated resistance to dengue virus infection and death at the cellular level. *PLoS One.* 2010; 5:e13398. [PubMed: 20976219]
- Ishmael N, Dunning Hotopp JC, Ioannidis P, Biber S, Sakamoto J, Siozios S, Nene V, Werren J, Bourtzis K, Bordenstein SR, Tettelin H. Extensive genomic diversity of closely related *Wolbachia* strains. *Microbiology.* 2009; 155:2211–2222. [PubMed: 19389774]
- Jeyaparakash A, Hoy MA. Long PCR improves *Wolbachia* DNA amplification: wsp sequences found in 76% of sixty-three arthropod species. *Insect Mol. Biol.* 2000; 9:393–405. [PubMed: 10971717]
- Karr TL, Yang W, Feder ME. Overcoming cytoplasmic incompatibility in *Drosophila*. *Proc. R. Soc. Lond. B.* 1998; 265:391–395.
- Klasson L, Walker T, Sebahia M, Sanders MJ, Quail MA, Lord A, Sanders S, Earl J, O'Neill SL, Thomson N, Sinkins SP, Parkhill J. Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. *Mol. Biol. Evol.* 2008; 25:1877–1887. [PubMed: 18550617]
- Knizewski L, Kinch LN, Grishin NV, Rychlewski L, Ginalski K. Realm of PD-(D/E)XK nuclease superfamily revisited: detection of novel families with modified transitive meta profile searches. *BMC Struct. Biol.* 2007; 7:40. doi:10.1186/1472-6807-7-40.
- Landmann F, Orsi GA, Loppin B, Sullivan W. *Wolbachia*-mediated cytoplasmic incompatibility is associated with impaired histone deposition in the male pronucleus. *PLoS Pathog.* 2009; 5:e1000343. [PubMed: 19300496]
- Laven, H. Speciation and evolution in *Culex pipiens*. In: Wright, J.; Pal, R., editors. *Genetics of insect vectors of disease*. Elsevier; Amsterdam: 1967. p. 251–275.
- Lee M, Bakir AA, Nguyen KN, Bachant J. The SUMO isopeptidase Ulp2p is required to prevent recombination-induced chromosome segregation lethality following DNA replication stress. *PLoS Genetics.* 2011; 7:e1001355. [PubMed: 21483811]
- Li S, Hochstrasser M. A new protease required for cell-cycle progression in yeast. *Nature.* 1999; 398:246–251. [PubMed: 10094048]
- Lobley A, Sadowski MI, Jones DT. pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. *Structural Bioinformatics.* 2009; 25:1761–1767. [PubMed: 19429599]

- Lorenzen MD, Gnirke A, Margolis J, Garnes J, Campbell M, Stuart JJ, Aggarwal R, Richards S, Park Y, Beeman RW. The maternal-effect, selfish genetic element *Medea* is associated with a composite Tc1 transposon. *Proc. Natl. Acad. Sci. USA*. 2008; 105:10085–10089. [PubMed: 18621706]
- Magnin M, Pasteur N, Raymond M. Multiple incompatibilities within populations of *Culex pipiens L.* in southern France. *Genetica*. 1987; 74:125–130. [PubMed: 3506532]
- McAleenan A, Cordon-Preciado V, Clemente-Blanco A, Liu I, Sen N, Leonard J, Jarmuz A, Aragon L. SUMOylation of the α -kleisin subunit of cohesion is required for DNA damage-induced cohesion. *Curr. Biol*. 2012; 22:1–12. [PubMed: 22197242]
- McMeniman CJ, Lane RV, Cass BN, Fong AWC, Sidhu M, Wang Y, O'Neill SL. Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti*. *Science*. 2009; 323:141–144. [PubMed: 19119237]
- Orth K, Xu Z, Mudgett MB, Bao ZQ, Palmer LE, Bliska JB, Mangel WF, Staskawicz B, Dixon JE. Disruption of signaling by *Yersinia* effector YopJ, a ubiquitin-like protein protease. *Science*. 2000; 290:1594–1597. [PubMed: 11090361]
- Presgraves DC. A genetic test of the mechanism of *Wolbachia*-induced cytoplasmic incompatibility in *Drosophila*. *Genetics*. 2000; 154:771–776. [PubMed: 10655228]
- Rances E, Voronin D, Tran-Van V, Mavingui P. Genetic and functional characterization of the type IV secretion system in *Wolbachia*. *J. Bacteriol*. 2008; 190:5020–5030. [PubMed: 18502862]
- Rankin DJ, Rocha EPC, Brown SP. What traits are carried on mobile genetic elements and why? *Heredity*. 2011; 106:1–10. [PubMed: 20332804]
- Rasgon JL, Ren X, Petridis M. Can *Anopheles gambiae* be infected with *Wolbachia pipientis*? Insights from an in vitro system. *Appl. Environ. Microbiol*. 2006; 72:7718–7722. [PubMed: 17028229]
- Salzberg SL, Puiu D, Sommer DD, Nene V, Lee NH. Genome sequence of the *Wolbachia* endosymbiont of *Culex quinquefasciatus* JHB. *J. Bacteriol*. 2008; 191:1725. [PubMed: 19114486]
- Serbus LR, Casper-Lindly C, Landmann F, Sullivan W. The genetics and cell biology of *Wolbachia*-host interactions. *Annu. Rev. Genet*. 2008; 42:683–707. [PubMed: 18713031]
- Shiny C, Krushna NSA, Babu S, Elango S, Manokaran G, Narayanan RB. Recombinant *Wolbachia* heat shock protein 60 (HSP60) mediated immune responses in patients with lymphatic filariasis. *Microbes Infect*. 2011 Doi:10.1016/j.micinf.2011.07.004.
- Siozios S, Ioannidis P, Klasson L, Andersson SGE, Braig HR, Bourtzis K. The diversity and evolution of *Wolbachia* ankyrin repeat domain genes. *PLOS One*. 2013; 8:e55390. [PubMed: 23390535]
- Taylor MJ, Voronin D, Johnston KL, Ford L. *Wolbachia* filarial interactions. *Cell Microbiol*. 2012 Doi: 10.1111/cmi.12084.
- Teixeira L, Ferreira A, Ashburner M. The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PloS Biol*. 2008; 6(12):e1000002. doi: 10.1371/journal.pbio.1000002.
- Vavre F, Charlat S. Making (good) use of *Wolbachia*: what the models say. *Curr. Opin. Microbiol*. 2012; 15:263–268. [PubMed: 22494817]
- Veneti Z, Clark ME, Zabalou S, Karr TL, Savakis C, Bourtzis K. Cytoplasmic incompatibility and sperm cyst infection in different *Drosophila-Wolbachia* associations. *Genetics*. 2003; 164:545–552. [PubMed: 12807775]
- Veneti Z, Clark ME, Karr TL, Savakis C, Bourtzis K. Heads or Tails: Host-parasite interactions in the *Drosophila-Wolbachia* system. *Appl. Environ. Microbiol*. 2004; 70:5366–5372. [PubMed: 15345422]
- Wimmer P, Schreiner S, Dobner T. Human pathogens and the host cell SUMOylation system. *J. Virol*. 2012; 86:642–654. [PubMed: 22072786]
- Yamaguchi Y, Park J, Inouye M. Toxin-antitoxin systems in bacteria and archaea. *Annu. Rev. Genet*. 2011; 45:61–79. [PubMed: 22060041]
- Zdobnov EM, Apweiler R. InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*. 2001; 17:847–848. [PubMed: 11590104]

HIGHLIGHTS

- Peptides encoded by *wPip_0282* occur in sperm from dissected mosquito spermathecae
- Homologs of *wPip_0282* are only present in CI-inducing *Wolbachia* strains
- *wPip_0282* and *wPip_0283* comprise a two gene operon that has duplicated and diverged
- *wPip_0283* encodes a C-terminal, eukaryotic SUMO protease domain

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

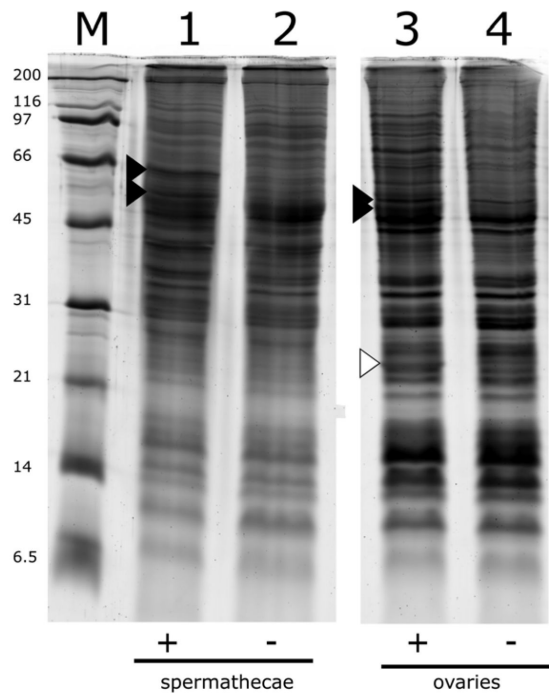


Figure 1. SDS-PAGE analysis of protein extracted from reproductive tissues from *Wolbachia*-infected and uninfected *Culex pipiens* mosquitoes. Lane M, molecular mass markers (kDa); lanes 1 and 2, 2400 dissected spermathecal lobes; lanes 3 and 4, 30 ovaries. Positive (+) and negative (-) symbols below 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 the mosquitoes had mated and that *Wolbachia* were absent from the spermathecal tissues.

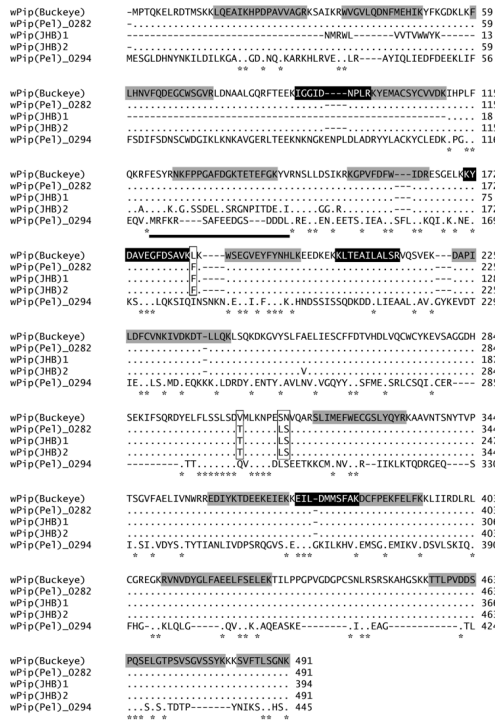


Figure 2. Total mass spectrometry detected peptide coverage of WPIP0282 and a sequence alignment of homologs from *Culex pipiens* strains of *Wolbachia*. Dots indicate amino acid identities with respect to wPip(Buckeye). Black shaded residues indicate peptides detected by mass spectrometry in both spermathecal and ovarian samples. Grey shaded residues indicate peptides detected only in ovarian samples. In wPip (Pe1)_0282 homologs, vertical boxes indicate regions of divergence with respect to wPip(Buckeye). The bold black underline indicates a region of divergence in wPip(JHB)2. wPip (Pe1)_0294 is a distant homolog of wPip (Pe1)_0282. Asterisks indicate fully conserved residues.

wPip(Pel) Operon Characteristics

A

wPip(Pel)	0282	0283	0294	0295	1291	1292
GI:190570728	GI:190570728	GI:190570729	GI:190570737	GI:190570738	GI:190571665	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 (identity%/similarity%)	to WPIP0282 100/100	to WPIP0283 100/100	to WPIP0282 34/53	to WPIP0283 23/42	to WPIP0283 (C-term) 99/99	to WPIP0283 (C-term) 38/54
E Value	0.0	0.0	2e-44	1e-24	0.0	4e-05

Homolog BLASTp Comparisons

B

Homology to:	0282	0283	0294	0295	1291	1292
wMel Operon	wMel_0631	wMel_0632	wMel_0631	wMel_0632	wMel_0632	wMel_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	wRI_005370	wRI_p05380	wRI_005370	wRI_p05380	wRI_p05380	wRI_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	wRI_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	776	375	727		
identity%/similarity%	35/49	30/47	28/45	31/51		
E Value	4e-45	7e-63	8e-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 wPip (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 wPip_0282; black arrows and unshaded text indicate homologs of wPip_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 wPip (Pel).

```

wPip(Buckeye)  MPTQKELRDTMSKQLQEAIKHPDPAVAVGRSAIKRWVGLQDNFMHAIKYFKGDKLFL 60
wRi_005370     ..IETKRQAEVL...DV...T.RDIA...L.....-ETYL.Y..L..D...E.. 56
wRi_010030     ..IETKRQAEVL...DV...T.RDIA...L.....-ETYL.Y..L..D...E.. 56
wMel_0631      ..IETKRQAEVL...DV...T.RDIA...L.....-ETYL.Y..L..D...E.. 56
wRi_006720     ..KKM.RHAAVLS...KSV.Q.T.SK.M.E.R...E.....-KTYIRQVE.L.D...Q.. 56
** * * * * *

wPip(Buckeye)  HNVFDEGCWGVRLDNAALGQRFTEEKIGGIDNPLRKYEMACSYCVVDKHPLFQKRFE 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.TE...E.K...PL.D...R...I...PL...Q... 116
* * * * *

wPip(Buckeye)  SYRNKFPFGAFDCKTETEF--GKYVNSLLDSIKRKGVPVDFWIDRESGELKKYD-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.SSEEL.DGKPATSNN...KSE..GYM.SQD...S..V.KK...F..HVS..T.. 176
*****

wPip(Buckeye)  FDSAVKLLWSEGVEFYFVNHLEEDKEK--KLTEATLALSRVQSEKDPALDFCVNKIVD 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...K.....RN.G. 231
wRi_006720     .KKAIE.....SL.N.KER.RER..I.D.VTI...SV.CDHNG.VT...LSKMS. 236
* * * * *

wPip(Buckeye)  --KDTLLQ--KLSQKDKGVYSLFAELIESCFDVTVDLVCWCYKEVSGAGDHSEKIFSQ 291
wRi_005370     --.....L.....G.....C.D... 287
wRi_010030     --.....L.....G.....C.D... 287
wMel_0631      --.....L.....G.....C.D... 287
wRi_006720     QA.NK.FKQSE..K.....SA..MQG...MQAILPMFKD.IL---ED..L.P 290
* * * * *

wPip(Buckeye)  RDYELFLSSLSDVMLKNPESNVQARSLIMEFWCCSLYQYR---KAAVNTSNYVPTSG 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)  VFAELIVNWRREDIYKTEDEEIEKKEI LDMMMSFAKD--CFPEKFLFKLIIRDRLCGR 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)  EGKRVNVDYGLFAEELFSELEKTLIPPGVGDGPCSNLRSRKAHGSKKTLPVD--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.--DFGAG-----D 438
* * * * *

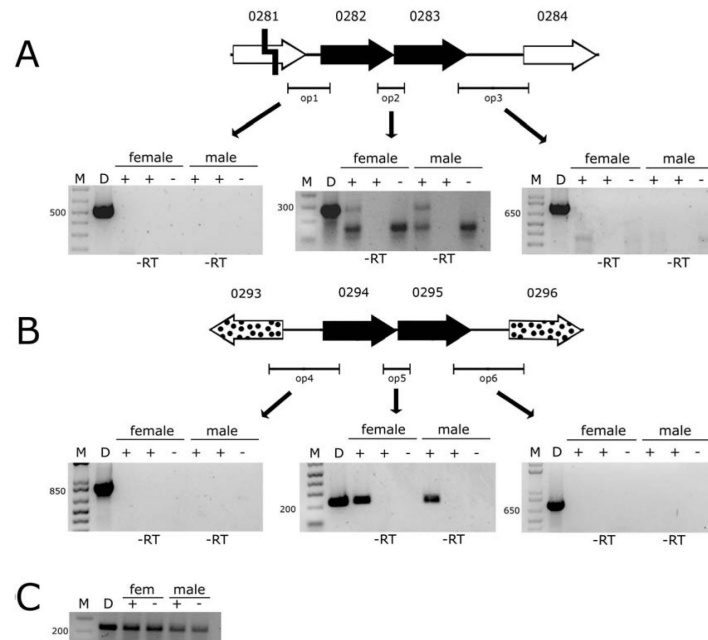
wPip(Buckeye)  POSELGTPSVSGVS-SYKKSQVFTLSGNK 491
wRi_005370     ...FEA...I.G.HKRRK----- 474
wRi_010030     ...FEA...I.G.HKRRK----- 474
wMel_0631      ...FEA...I.G.HKRRK----- 474
wRi_006720     ...T..STE..SF.GRQ----- 456
***

```

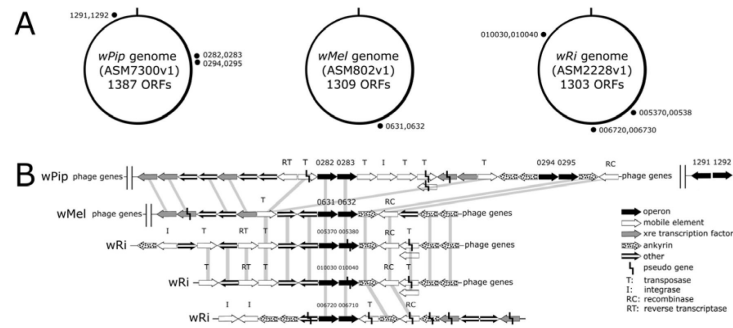
Figure 4. Protein alignments of other WPA0282 homologs within *wMel* and *wRi*. Dots indicate identities with respect to *wPip*(Buckeye) and asterisks indicate fully conserved residues. The bold black underline highlights the region of divergence corresponding to that shown for *wPip*(JHB)2 in Fig 2.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



Figure 5. Protein alignments of the second gene in the operon, *wPip0283*, and its homologs in *wPip(Pel)*. Dots indicate identities with respect to *wPip(Pel)_0283*. Grey shaded residues show the C-terminally truncated homolog, *wPip0295*; its terminus is indicated by a grey circle. Black shaded residues show N-terminally truncated homolog, *wPip1291*; 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, *wPip1292*; 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 *wPip_1291* begins downstream of the C-terminus of *wPip_0295*.

**Figure 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.

**Figure 7.**

Operon structure and synteny within the three *Wolbachia* strains, *wPip* endosymbiont of *Culex quinquefasciatus*, *wMel* endosymbiont of *Drosophila melanogaster*, and *wRi* endosymbiont of *Drosophila simulans*. **A.** Genomes of the three strains with locations of the duplicated operon elements labeled as black dots. In *wPip*, the operon has been fully duplicated one time with parts of 0283 also being duplicated another two times to 1291 and 1292. In *wMel* the operon only occurs once. In *wRi* the operon has been duplicated three times throughout the genome. **B.** Expanded genome regions including the respective operon elements in the various species. Genes *wRi*_005380 and *wRi*_010040 are labeled as pseudo genes within the genome, but still encode a large open reading frame. Genes *wPip*_0283, *wPip*_1291, *wMel*_0632, *wRi*_p005380, and *wRi*_p010040 all have a conserved eukaryotic C48 sumo protease protein domain. Gene sizes are not to scale. Vertical shading indicates synteny among genome regions.