CrossMark

ORIGINAL PAPER

Mosaic composition of *rib*A and *wsp*B genes flanking the *vir*B8-D4 operon in the *Wolbachia* supergroup B-strain, *w*Str

Gerald D. Baldridge[1] · Yang Grace Li[1] · Bruce A. Witthuhn[2] · LeeAnn Higgins[2] · Todd W. Markowski[2] · Abigail S. Baldridge[3] · Ann M. Fallon[1]

Abstract The obligate intracellular bacterium, *Wolbachia pipientis* (*Rickettsiales*), is a widespread, vertically transmitted endosymbiont of filarial nematodes and arthropods. In insects, *Wolbachia* modifies reproduction, and in mosquitoes, infection interferes with replication of arboviruses, bacteria and plasmodia. Development of *Wolbachia* as a tool to control pest insects will be facilitated by an understanding of molecular events that underlie genetic exchange between *Wolbachia* strains. Here, we used nucleotide sequence, transcriptional and proteomic analyses to evaluate expression levels and establish the mosaic nature of genes flanking the T4SS *vir*B8-D4 operon from *w*Str, a supergroup B-strain from a planthopper (Hemiptera) that maintains a robust, persistent infection in an *Aedes albopictus* mosquito cell line. Based on protein abundance, *rib*A, which contains promoter elements at the 5′-end of the operon, is weakly expressed. The 3′-end of the operon encodes an intact *wsp*B, which encodes an outer membrane protein and is co-transcribed with the *vir* genes. WspB and vir proteins are expressed at similar, above average abundance levels. In *w*Str, both *rib*A and *wsp*B are mosaics of conserved sequence motifs from *Wolbachia* supergroup A- and B-strains, and *wsp*B is nearly identical to its homolog from *w*CobU4-2, an A-strain from weevils (Coleoptera). We describe conserved repeated sequence elements that map within or near pseudogene lesions and transitions between A- and B-strain motifs. These studies contribute to ongoing efforts to explore interactions between *Wolbachia* and its host cell in an in vitro system.

Keywords *Wolbachia* · LC–MS/MS · Proteomics · Mosaic genes · T4SS · RibA · RibB · WspB

✉ Ann M. Fallon
fallo002@umn.edu

[1] Department of Entomology, University of Minnesota, 1980 Folwell Ave., St. Paul, MN 55108, USA

[2] Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN 55455, USA

[3] Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA

Introduction

Wolbachia pipientis (*Rickettsiales*; *Alphaproteobacteria*) is an obligate intracellular bacterium that infects filarial nematodes and a wide range of arthropods including ≥60 % of insects and ≈35 % of isopod crustaceans, but does not infect vertebrates (Hilgenboecker et al. 2008). *Wolbachia* is considered to be a single species classified into clades by multilocus sequence typing and designated as supergroups A to N (Baldo et al. 2006b; Comandatore et al. 2013; Lo et al. 2007). The C- and D-strains that infect filarial worms have phylogenies concordant with those of nematode hosts, consistent with strict vertical transmission as obligate mutualists (Comandatore et al. 2013; Dedeine et al. 2003; Li and Carlow 2012; Strubing et al. 2010; Taylor et al. 2005; Wu et al. 2004). Although arthropod-associated A- and B-strains may provide subtle fitness benefits to hosts (Zug and Hammerstein 2014), they are best known as reproductive parasites, causing phenotypes that maintain or increase *Wolbachia* infection frequencies, including

feminization, parthenogenesis, and cytoplasmic incompatibility (Saridaki and Bourtzis 2010; Werren et al. 2008). Interference with host immune mechanisms and replication of arboviruses, bacteria and malarial plasmodia (Kambris et al. 2009; Pan et al. 2012; Zug and Hammerstein 2014) has encouraged efforts to exploit *Wolbachia* for biocontrol of arthropod vectors of vertebrate pathogens and/or crop pests (Bourtzis 2008; Rio et al. 2004; Sinkins and Gould 2006; Zabalou et al. 2004). An understanding of molecular differences between A- and B-strains, and how they have been influenced by horizontal transmission and genetic exchange (Newton and Bordenstein 2011; Schuler et al. 2013; Werren et al. 2008; Zug and Hammerstein 2014) will facilitate manipulation of *Wolbachia*.

Wolbachia's interaction with host cells likely involves the type IV secretion system (T4SS), a macromolecular complex that transports DNA, nucleoproteins and "effector" proteins across the microbial cell envelope into the host cell, where they mediate intracellular interactions (Alvarez-Martinez and Christie 2009; Zechner et al. 2012). Homologs of all genes except *vir*B5 of *Agrobacterium tumefaciens* T4SS have been identified in *Wolbachia* and other members of the *Rickettsiales* (Gillespie et al. 2009, 2010), including *Anaplasma*, *Ehrlichia*, *Neorickettsia*, *Orientia* and *Rickettsia.* Among sequenced *Wolbachia* genomes, T4SS genes are organized in two operons: *vir*B3-B6 containing *vir*B3, *vir*B4 and four *vir*B6 paralogs and *vir*B8-D4 containing *vir*B8, *vir*B9, *vir*B10, *vir*B11, *vir*D4 and, in some genomes, the *wsp*B paralog of the *wsp*A major surface antigen (Pichon et al. 2009; Rances et al. 2008). In the supergroup B-strain *w*Pip from *Culex pipiens* mosquitoes, *wsp*B is disrupted by a transposon and is presumably inactive (Sanogo et al 2007). T4SS effector proteins that manipulate host cells have been identified from *Anaplasma* and *Ehrlichia* (Liu et al. 2012; Lockwood et al. 2011; Niu et al. 2010), and *Wolbachia* express both *vir* operons in ovaries of arthropod hosts, wherein T4SS effectors are suspected to play a role in cytoplasmic incompatibility and other reproductive distortions (Masui et al. 2000; Rances et al. 2008; Wu et al. 2004). Although WspA and WspB are likely components of the *Wolbachia* outer membrane, their functions remain unknown. In the case of *w*Bm, WspB is excreted/secreted into filarial host cells (Bennuru et al. 2009) and co-localizes with the Bm1_46455 host protein in tissues that include embryonic nuclei (Melnikow et al. 2011). WspB is therefore itself a candidate T4SS effector that may play a role in reproductive manipulation of the host.

The *Wolbachia* strain *w*Str in supergroup B causes strong cytoplasmic incompatibility in the planthopper, *Laodelphax striatellus* (Noda et al. 2001a), and in addition maintains a robust, persistent infection in a clonal *Aedes albopictus* mosquito cell line, C/*w*Str1 (Fallon et al. 2013; Noda et al. 2002). Because in vitro studies with *w*Str provide advantages of scale and ease of manipulation for exploring mechanisms that may facilitate transformation and genetic manipulation of *Wolbachia*, we have undertaken proteomics-based studies that provide strong support for expression of T4SS machinery in cell culture. Here, we report the sequence of the *vir*B8-D4 operon, including flanking genes *rib*A, upstream of *vir*B8, and *wsp*B downstream of *vir*D4. We show that *wsp*B is intact, describe protein structure predicted from the deduced WspB sequence, and verify co-transcription of *wsp*B with upstream *vir* genes. Relative abundance levels of WspB and the VirB8-D4 proteins in *w*Str are well above average, while RibA is among the least abundant of MS-detected proteins. In *w*Str, *rib*A and *wsp*B are mosaics of sequence motifs that are differentially conserved in supergroup A- (WOL-A) and B- (WOL-B) strains, and they contain conserved 8-bp repeat elements that may be associated with genetic exchange. Finally, we discuss implications for functional integration of the *Wolbachia* T4SS with WspB and with the riboflavin biosynthesis pathway enzymes GTP cyclohydrolase II (RibA) and dihydroxybutanone phosphate synthase (RibB).

Materials and methods

Cultivation of cells

Aedes albopictus C7-10 and C/*w*Str1 cells were maintained in Eagle's minimal medium supplemented with 5 % fetal bovine serum at 28–30 °C in a 5 % CO_2 atmosphere (Fallon et al. 2013; Shih et al. 1998). Cells were harvested during exponential growth, under conditions favoring maximal recovery of *Wolbachia* (Baldridge et al. 2014).

Polymerase chain reaction, cloning and DNA sequencing

The polymerase chain reaction (PCR) was used to amplify *w*Str genes from DNA extracts prepared from *Wolbachia* enriched by fractionation of C/*w*Str1 cells on sucrose density gradients and recovered from the interface between 50 and 60 % sucrose (Baldridge et al. 2014). Template DNA was used to obtain 21 PCR products using a panel of 31 primers (Table S1), GoTaq™ DNA polymerase (Promega, Madison, WI), and a Techne TC-312 cycler (Staffordshire, UK). Cycle parameters were: 1 cycle at 94 °C for 2 min, 35 cycles at 94 °C for 35 s, 53 °C for 35 s, 72 °C for 1 min, followed by 1 cycle at 72 °C for 5 min. Extension time was increased to 2 min for products ≥1000 bp. PCR products were cloned in the pCR4-TOPO vector with the TOPO-TA Cloning Kit for Sequencing (Life Technologies, Grand Island, NY), and two or more clones each were sequenced

Table 1 MS-detected peptides from *w*Str proteins encoded by *rib*A, *rib*B and the *vir*B8-D4 operon

Protein	[a]kDa	[b]Pep(1)	[b]Pep(2)	[b]Pep(T)	[c]Cov.	[d]RAL	[e]SR
RibA	41	2	2	2	6	0.5	−2.30
RibB	24	7	12	12	89	7.0	1.20
VirB8	26	9	10	10	58	5.0	0.59
VirB9	31	10	8	10	45	6.2	0.84
VirB10	54	14	16	18	53	8.8	0.94
VirB11	37	12	14	14	42	7.0	0.82
VirD4	77	12	14	14	26	6.2	0.45
WspB	31	2[f]	11	50	7.2	1.08	

[a] Protein mass in kilodaltons. [b] Number of 95 % confidence unique peptides; (1) designates original search [7]; (2) designates a refined search in which the database included peptides based on the present *w*Str nucleotide sequence data; (T) combined total peptides from both searches. [c] Percent protein sequence coverage represented by detected peptides. [d] Mean number of peptides from four independent MS data sets. [e] Studentized residual based on the modified univariable model of the refined search (Table S3, column R); SR value 0 indicates average abundance protein, 0–1 above average, 1–2 abundant and >2 highly abundant. Values below 0 indicate lower than average abundance. [f] A 94 % confidence peptide indicated in Fig. 1A did not meet the threshold for proteome inclusion in the original search. For VirB10, one originally detected peptide was absent from the refined search

at the University of Minnesota BioMedical Genomics Center.

Reverse transcriptase polymerase chain reaction

Total RNA was purified from *A. albopictus* C7-10 and C/*w*Str1 cells using the PureLink RNA Mini Kit (Life Technologies) and treated with DNase I (RNase-free; Life Technologies) followed by heat inactivation, as suggested by the manufacturer. RT-PCR was executed with primers *vir*D4$_{F1764–1784}$ and *wsp*B$_{R152–172}$ (Table S1) using the RNA PCR Core Kit (Life Technologies) as suggested by the manufacturer with the exception that synthesized cDNA was treated with DNase-inactivated RNaseA before the final PCR reaction. The PCR reaction included 1 cycle at 95 °C for 4 min, 35 cycles at 95 °C for 35 s, 56 °C for 40 s, 72 °C for 40 s, followed by 1 cycle at 72 °C for 3 min. Reaction products were electrophoresed on 1 % agarose gels, cloned, and sequenced as above.

Sequence alignments and protein structure prediction

DNA and protein sequence alignments were executed with the Clustal Omega program (Sievers et al. 2011). Alignments were edited by visual inspection and modified in Microsoft Word. WspB protein structure predictions were obtained using tools available at www.predictprotein.org, including the PROFtmb program (Dell et al. 2010) for prediction of bacterial transmembrane beta barrels (Bigelow et al. 2004) and per-residue prediction of up-strand, down-strand, periplasmic loop and outer loop positions of residues. The PROFisis program (Ofran and Rost 2006) was used to predict WspB amino acid residues that are

potentially involved in protein–protein interactions. Trees were produced using PAUP* version 4 (Swofford 2002). Amino acids were aligned with Clustal W, using pairwise alignment parameters of 25/0.5 and multiple alignment parameters of 10/0.2 for gap opening and gap extension, respectively. The protein weight matrix was set to Gonnet. The alignment was saved as a nexus file and loaded into PAUP*, and the trees were created using a heuristic search with the criterion set to parsimony. Bootstrap 50 % majority-rule consensus trees are based on 1000 replicates, with *w*Bm (WOL-D) as the outgroup.

Mass spectrometry, peptide detection, protein identification and statistical analysis

Mass spectrometry data, generated using LC–MS/MS on LTQ and Orbitrap Velos mass spectrometers as four data sets, were described previously (Baldridge et al. 2014). The MS search database was modified to include deduced ORFs from *w*Str sequence data described herein. All tests of association were performed with SAS version 9.3 (Cary, NC; http://www.sas.com/en_us/home.html/).

Results

Structure of the *w*Str *vir*B4-D8 operon

The robust, persistent infection of *A. albopictus* mosquito cell line, C/*w*Str1 with [B]*w*Str (in the text below, strain designations are denoted by superscripts), isolated from the planthopper *L. striatellus*, provides an in vitro model to identify proteins that modulate the host–microbe
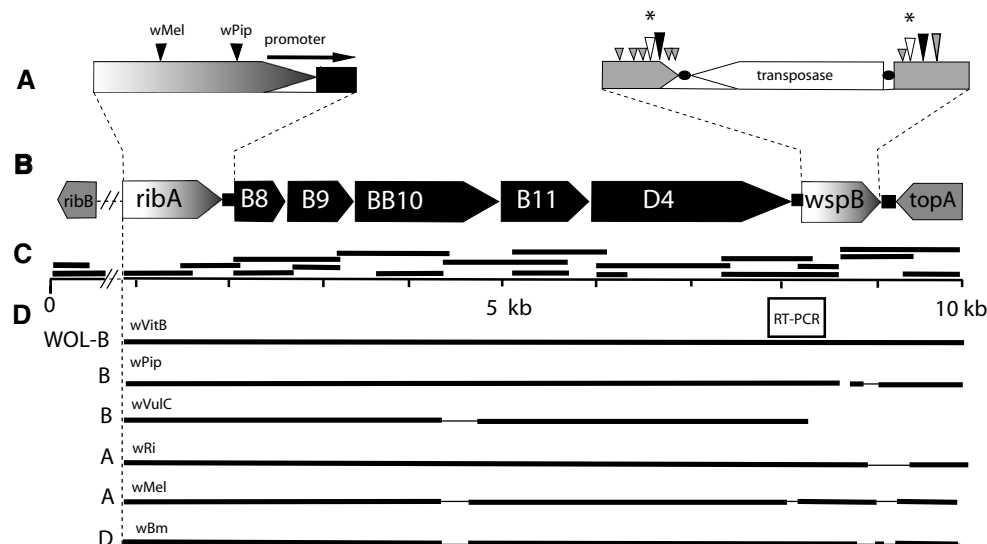
Fig. 1 Schematic map of the *Wolbachia* T4SS *vir*B8-D4 operon and cloning strategy for the *rib*A to *top*A sequence from [B]*w*Str. **a** *Left* expanded view of the [B]*w*Str *rib*A ORF depicted as an *arrow* showing the direction of transcription. *Black horizontal arrow* indicates a putative promoter that extends into an intergenic spacer (*black rectangle*). *Black arrowheads* indicate positions of MS-detected unique peptides (95 % confidence). Gradient shading from *white* to *black* designates 5′-sequence identity resembling WOL-A transitioning to 3′-sequence more closely resembling WOL-B-strains. **a** *Right* expanded view of the interrupted *wsp*B homolog in [B]*w*Pip. *Black ellipses* indicate positions of IS256 inverted repeat elements flanking a 1.2-kb insertion encoding a MULE domain superfamily transposase (gi|190571636; pfam10551) on the opposite strand (indicated by the direction of the *open arrow*); flanking *gray shading* indicates *wsp*B. *Tall vertical black and gray arrowheads* indicate positions of unique peptides (95 and 94 % confidence, respectively) identified in the original MS data search. *Small gray arrows* indicate 95 % confidence peptides matched in a refined data set (including the [B]*w*Str sequence described here)

that are conserved in WOL-B-strains, and *open arrowheads with stars* indicate peptides unique to [B]*w*Str. **b** Schematic depiction of the *Wolbachia vir*B8-D4 operon and flanking genes with *arrows* designating the direction of transcription. *Vir* genes are designated in *white font on a black background*; *black squares* indicate intergenic spacers. *Gradient shading* indicates mosaic structure of an intact wspB in [B]*w*Str. **c** *Filled lines* above the 10-kb scale marker represent cloned PCR amplification products (see Table S1 for primers) that were sequenced and assembled into the [B]*w*Str *rib*B and *rib*A–*top*A consensus sequence. The *double slash symbols at left* indicate that *rib*B is not contiguous with downstream genes. The *open box* indicates the RT-PCR amplification product from Fig. 2. **d** BLASTn alignment of the 9133-bp [B]*w*Str *rib*A–*top*A sequence to corresponding sequences in [B]*w*VitB [B]*w*Pip, [B]*w*VulC, [A]*w*Ri, [A]*w*Mel and [D]*w*Bm genomes. *Dark filled lines* indicate sequence identity >70 %; *light lines* indicate low sequence identity, and the open space in [B]*w*Pip represents an alignment gap

interaction. A potential role for the T4SS is supported by strong representation of peptides from VirB8, VirB9, VirB10, VirB11, VirD4 (Table 1) and associated proteins in the [B]*w*Str proteome (Baldridge et al. 2014). Despite its emergence as a useful strain that grows well in vitro, the [B]*w*Str genome is not yet available. In *Wolbachia* strains for which genome annotation is available, gene order within the *vir*B8-D4 operon is conserved. Based on transcriptional analyses in the related genera, *Anaplasma* and *Ehrlichia* (Pichon et al. 2009), the promoter likely maps within the 3′-end of *rib*A extending into the intergenic spacer (Fig. 1a, black horizontal arrow at left) and is followed by five consecutive *vir* genes (Fig. 1b). In [B]*w*Pip from *Culex pipiens* mosquitoes, *wsp*B is disrupted by insertion of an IS256 element that encodes a transposase on the opposite strand (Fig. 1a, at right; Sanogo et al. 2007). Because VirB8-D4 proteins were highly similar to homologs from [B]*w*Pip (Baldridge et al. 2014), we evaluated *wsp*B in [B]*w*Str and its potential expression as a *vir*B8-D4 operon member, as

is the case in [A]*w*Mel and [A]*w*Ri from *Drosophila spp.* and [A]*w*Atab 3 from the wasp *Asobara tabida* (Rances et al. 2008; Wu et al. 2004). In the original proteomic analysis, three WspB peptides (Fig. 1a, tall black and gray arrows represent 95 and 94 % confidence peptides, respectively) mapped proximal and distal to the transposon insertion in [B]*w*Pip, while the absence of peptides corresponding to the transposon suggested that *wsp*B is intact in [B]*w*Str.

Nucleotide and deduced amino acid sequence comparisons

To examine the *vir*B4-D4 operon in [B]*w*Str, we sequenced overlapping PCR products from 20 primer pairs (Table S1) spanning 9.1 kb beginning 43 bp downstream of the 5′-end of *rib*A in other *Wolbachia* strains and ending within *top*A encoded immediately downstream of the operon on the opposite strand (Fig. 1b, c). With the notable exception of the [B]*w*Pip transposon, the nucleotide sequence aligned most

Table 2 Pairwise nucleotide and amino acid comparisons

Gene	[B]wPiP		[B]wVitB		[B]wNo		[B]wTai		[B]wVulC		[A]wMel		[A]wRi	
	N	AA	N	AA	N	AA	N	AA	N	AA	N	AA	N	AA
*rib*A[a]	94	89	94	89	93	88	94	90	93	92	93	91	92	89
*vir*B8	*99*	*100*	*99*	*100*	*99*	*99*	*99*	*100*	94	94	88	86	88	87
*vir*B9	*99*	*99*	*99*	*98*	*98*	*97*	*97*	*97*	94	93	91	89	91	89
*vir*B10	*99*	*99*	*99*	*98*	90	86	*98*	96	88	74	87	74	88	85
*vir*B11	*99*	*99*	*97*	*99*	96	*98*	*97*	*99*	90	93	89	95	89	95
*vir*D4	*99*	*99*	*99*	*99*	*99*	*99*	*99*	*99*	94	*97*	89	92	89	93
*wsp*B	56	xx	*98*	96	85	68	–	–	–	–	85	70	85	70
*top*A[a]	*99*	*100*	*99*	*100*	*99*	*99*	–	–	–	–	88	87	87	86
*rib*B[a]	81	80	–	–	*97*	96	–	–	–	–	90	91	79	78

Gene	[A]wAna		[A]wKue		[A]wAtab3		[F]wCle		[D]wBm		[C]wOo		[C]wOv	
	N	AA	N	AA	N	AA	N	AA	N	AA	N	AA	N	AA
*rib*A[a]	91	88	93	91	–	–	84	81	83	80	82	74	82	75
*vir*B8	88	87	88	86	88	88	85	83	85	81	83	81	84	82
*vir*B9	91	89	91	89	91	89	84	84	84	84	82	76	81	76
*vir*B10	88	84	87	74	87	73	80	71	84	70	76	64	74	64
*vir*B11	89	95	89	95	89	95	89	95	88	94	86	89	87	89
*vir*D4	89	93	89	93	88	92	87	92	87	87	86	91	88	94
*wsp*B	83	68	85	70	85	70	72	xx	73	61	72	49	71	49
*top*A[a]	86	85	–	–	–	–	88	92	86	88	84	83	84	88
*rib*B[a]	80	88	–	–	–	–	87	87	86	87	85	xx	85	xx

Wolbachia strains from supergroups A, B, C, D and F are indicated by superscripts, with percentages of nucleotide (N) and amino acid (AA) sequence identities to [B]wStr. Dashes indicate sequences not available, and xx indicates pseudogenes; GenBank Accession numbers are given in Table S2

[a] Partial gene and protein sequences: *rib*A 1040 bp, *rib*B 592 bp; *top*A 825 bp. Host associations: *w*Pip, *Culex pipiens*—mosquito; *w*VitB, *Nasonia vitripennis*—wasp; *w*Tai, *Teleogryllus taiwanensis*—cricket; *w*VulC, *Armadillidium vulgare*—isopod; *w*Mel, *w*Ri, *w*Ana, *w*No, *Drosophila* spp.—fruit fly; *w*Kue, *Ephestia kuehniella*—moth; *w*Atab 3 *Asobara tabida*—wasp; *w*Bm, *w*Oo and *w*Ov from filarial nematodes *Brugia malayi*, *Onchocerca ochengi* and *O. volvulus,* respectively. In the comparison, values of 97 % or greater are shown in italics

closely to homologous sequences from [B]wVitB and [B]wPip. In addition, we noted variability in an ~0.3-kb region of *vir*B10 in [B]wStr that was conserved in [B]wVitB, [B]wPip and [A]wRi, but not in [B]wVulC, [A]wMel and [D]wBm (Fig. 1d; see Table S2 for GenBank Accessions).

Pairwise sequence comparisons of the *vir*B8-D4 operon from [B]wStr to homologs from *Wolbachia* supergroup A, B, C, D and F strains (Table 2) confirm that *vir*B10, with nucleotide identities ranging from 74–99 %, is the least conserved of the five *vir* genes, and we note that Klasson et al. (2009) attributed divergence of *vir*B10 in [A]wMel and [A]wRi to genetic exchange with a WOL-B-strain. Collectively and as individuals, the *vir* genes from [B]wStr have the highest nucleotide identities (~99 %) with [B]wVitB and [B]wPip. Identities with five A-strains are lower (range 87–91 %), lower yet (range 80–89 %) with the F-strain, [F]wCle and fall to a range of 74–88 % with three nematode-associated strains, [D]wBm, [C]wOo and [C]wOv. At the 5′-end of the operon, *rib*A was distinct, with approximately equivalent nucleotide identity with homologs from A- and B-strains (range 91–94 %), while the partial sequence of *top*A downstream of the operon had a conservation pattern similar to that of the *vir* genes. In some comparisons, *vir*B8, *vir*B11, *vir*D4 and *top*A amino acid identities exceed nucleotide identities. Although *rib*B is not physically adjacent to the *vir*B8-D4 operon in annotated *Wolbachia* genomes, *rib*B from [B]wStr is most similar to homologs from [B]wNo (97 % nucleotide identity) and [A]wMel (90 %), but was exceptional because identities with three other insect-associated A- and B-strains (~80 %) were lower than with F-, C- and D-strains (range 85–87 %). Consistent with earlier proteomic data (Baldridge et al. 2014), in all comparisons that discriminate between A- and B-strains, [B]wStr resembled WOL-B, while variability in *rib*A and *wsp*B flanking the *vir*B8-D4 genes exceeded that of the *vir* genes themselves.

Expression and relative abundances of the [B]wStr virB4-D8 proteins

To refine an earlier original proteomic analysis (Baldridge et al. 2014), we incorporated the PCR-amplified [B]wStr sequences described here to the database for peptide identification [Table 1, see column labeled Pep(2)]. Statistical analysis indicated that in a univariable model, protein molecular weight was weakly ($r^2 = 0.2221$) but significantly ($p < 0.0001$) associated with peptide count: log(*peptides*) = $-0.40247 + 0.4953 \times$ log(*MW*). Estimations of protein relative abundance levels (RAL) based on peptide counts were therefore normalized to protein length using studentized residuals (SR), a measure of deviance from expected values adjusted for estimated SD from the mean. All peptide data and SR values in the univariable and multivariable models of the original and refined searches are detailed in Table S3.

In the refined search, we identified eight new peptides from Vir proteins [Table 1, compare columns labeled Pep(2) to Pep(1)], including three from the most divergent VirB10. In aggregate, the five Vir proteins had a mean (SD) SR of 0.73 (0.2) and are expressed at above average abundance. We identified five new peptides from RibB, but none from RibA (Table 1). RibB has an SR of 1.2 and is an abundant protein, while RibA has an SR of -2.3 and is among the least abundant of MS-detected proteins. Nine new peptides from the highly divergent WspB (see below) generated an SR of 1.08, slightly above the threshold (>1.0) for an abundant protein and roughly equivalent to SR values (range 1–1.17) of housekeeping proteins such as isocitrate dehydrogenase, ftsZ, ATPsynthase F0F1 α subunit, and ribosomal proteins S2, S9, L3, L7/L12 and L14 (Table S3). In comparison, WspA with an SR of 2.17 (Table S3, entry 63) ranked as highly abundant, and the most abundant protein in the proteome was the GroEL chaperone (entry 586), with an SR of 3.66.

Reverse transcriptase PCR confirms co-transcription of *wsp*B with vir genes

Similar SR values for WspB, relative to VirB8-D4, were consistent with evidence that *wsp*B is co-transcribed with *vir*B8-D4 in [A]wMel, [A]wRi and [A]wAtab 3 (Rances et al. 2008; Wu et al. 2004). We used RT-PCR with RNA template verified by PCR to be free of DNA contamination (Fig. 2b, lanes 2 and 3) to amplify a 528-bp product that was produced in reactions containing RNA from C/wStr1 cells (Fig. 2a, lane 4), but not in negative control reactions (lanes 1 and 2) or those with RNA from C7-10 cells (lane 3). Its sequence matched the expected [B]wStr genomic sequence (Fig. 1c, RT-PCR box at right), confirming that in [B]wStr, *wsp*B is a member of the *vir*B8-D4 operon.
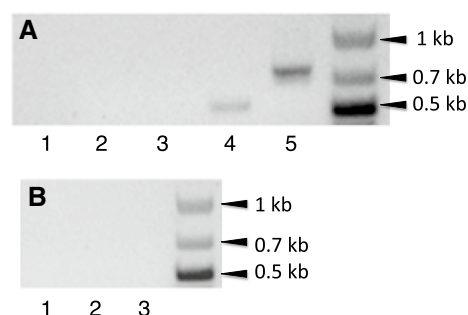

Fig. 2 Reverse transcriptase PCR (RT-PCR) analysis shows co-transcription of *wsp*B with *vir*D4. **a** *Lanes 1 and 2* RT-PCR negative controls with no RNA or with no reverse transcriptase, respectively. *Lanes 3 and 4* RT-PCR of RNA from uninfected C7-10 and infected C/wStr1 cells, respectively, with *vir*D4 forward and *wsp*B reverse primers. *Lane 5* RT-PCR positive control with C/wStr1 RNA and *Wolbachia* primers S12F/S7R, which amplify portions of a ribosomal protein operon described previously (Fallon 2008). **b** *Lane 1* PCR negative control with no Taq enzyme. *Lanes 2 and 3* negative control lacking RT, with RNA from uninfected C7-10 and infected C/wStr1 cells, respectively

In [B]wStr, *rib*A is a mosaic of conserved WOL-A and WOL-B sequence motifs

The *rib*A nucleotide sequence has been shown to contain regulatory elements for expression of the T4SS operon in *Anaplasma* and *Ehrlichia* (Ohashi et al. 2002; Pichon et al. 2009). In contrast to highest homologies of [B]wStr *vir*B8-D4 genes to WOL-B-strains, *rib*A sequence identities showed little difference between WOL-A and -B homologs (Table 2), but the two MS-detected peptides corresponded to [A]wMel and [B]wPip homologs, respectively (Fig. 1a). Alignment of amino acids from 10 RibA homologs (Fig. 3; WOL-A and WOL-B-strains are identified at left in red and blue, respectively) suggested that [B]wStr RibA is a two-part mosaic, each containing a protein functional domain.

The amino terminal 150 residues in [B]wStr RibA (Fig. 3) include a short dihydroxybutanone phosphate synthase domain and the first detected peptide (residues 94–104). This portion of [B]wStr RibA matched sequences from the four A-strains and a single B-strain, [B]wVulC, at 29 of 36 variable amino acids (shown in red), while only three (4, 39 and 168 in blue) matched the other three B-strains and four (in green) were unique. In contrast, the C-terminal 151–347 residues, encompassing the second peptide (residues 250–258) within a GTP cyclohydrolase domain, included a single amino acid unique to [B]wStr, while 23 (in blue) uniformly matched B-strains except [B]wVulC, which continued to resemble the A-strains until residue 239. Among the four A-strains, the [B]wRi homolog is most similar throughout the alignment to the B-strains, but within residues 129–150 immediately preceding the cyclohydrolase domain, it closely matched [B]wTai, [B]wPip and [B]wVitB, while [B]wStr

Fig. 3 Amino acid sequence alignment of RibA homologs from [B]wStr and *Wolbachia* supergroups A (*red*), B (*blue*) and D (*black*) respectively. *Asterisks below the alignment* indicate universally conserved residues. Unique residues are in *green font*. Residues conserved in [B]wStr and a majority of B-strains are in *dark blue, bold font*, while those in *dark red, bold font* are conserved with a majority of A-strains. Residues conserved in two to four strains are in *light blue, orange* or *orange bold font*. Residues highlighted in *gray* correspond to 95 % confidence peptides detected by LC–MS/MS. The dihydroxybutanone phosphate synthase (RibB) and GTP cyclohydrolase II domains (RibA) are indicated above the alignment within *greater than less than* symbols. *Bold underlined* residues in [A]wMel and [B]wStr indicate conserved active site amino acids, including critical cysteine residues. *Double underlined* residues indicate amino acids involved in the dimerization interface. See Tables 2 and S2 for host associations and GenBank Accessions. The PCR-amplified [B]wStr sequence does not encode the N-terminal amino acids; position 1 corresponds to the 15th amino acid

```
        1> DHBP synthase domain                <                          60
wKue    ISEIRRGRPI VIYDE.SNYL LFAAAEALER DLFNQYKLTS SNVYVTLTSS KVKYISQNKE
wMel    ISEIRRGRPI VIYDE.SNYL LFAAAEALER DLFNQYKLTS SNVYVTLTSS KVKYISQNKE
wHa     ISEIRRGRPI VIYDE.SNYL LFAAAEALER DLFNQYKLTS SNVYVTLTSS KVKYISQNKE
wRi     ISEIRRGRPI VIYDE.SNYL LFAAAEALER DLFNQYKLIS SNVYVTLTSS KVKYISQNKE
wVulC   ISEIRSGRPI VIYDE.SNYL LFAAVEALER DLFNQYKLIS SNVYVTLTSS KVKYISQNKE
wStr    ISEVRRGRPI VIYDE.SNYL LFAAAEVLER DLFNQYKLIS SNVYVTLTSS NVKYISQNKE
wTai    ISEVRRGLPI LIYDDKNNYL LFAAAETLEK NLFSQYKLIS GNVYVTLTAS KVKYICQSKE
wPip    ISEVRRGLPI LIYDDENNYL LLAAAETLEK NLFSQYKLIS GNVYVTLTAS KVKYICQSKE
wVitB   ISEVRRGLPI LIYDDENNYL LLAAAETLEK NLFSQYKLIS GNVYVTLTAS KVKYICQSKE
wBm     ISEIRRGLPI IIYDK.SNYL LVAAAETLEK DLFNQYGLIS GKIYVILPSS KVTCISQNVE
        *** * * **  ***    *** * **** **  ** ** * *   ** *  *  *  * * *

        61                                                              120
wKue    HNSKRLLVNN FDELLYLINC SKEDCIKELQ CSKTIDECAI ALLKFSELLP YALVADMTFE
wMel    HNSKRLLVNN FDELLYLINC SKEDCIKELQ CSKTIDECAI ALLKFSELLP YALVADMTFE
wHa     HNSKRLLVNN FDELLYLINC SKEDCIKELQ CSKTIDECAI ALLKFSELLP YALVADMTFE
wRi     HNSKRLLVNN FDELLHLINC SKEDCIKELQ CSKTIDECAI ALLKFSELLP YALVADMTFE
wVulC   HNSKRLLVNN FDELLYLINC SKEDCIKELQ CSKTIDECAI ALLKFSELLP YALVADMTFE
wStr    HNSKRLLVNN FDELLYLINC SKEDCMKELQ CSKTIDECAI ALLKFSELLP YALVADMTFE
wTai    HSSKRLLVNN FDELLHLIDC SKEDHIKELQ CSKTIDEYAI ALLKFSELLP YALVADMTFE
wPip    HSSKRLLISN FDELLHLINC SKEDHIKELQ CSKTIDEYAI ALLKFSELLP YALVADMTFE
wVitB   HSSKRLLISN FDELLHLINC SKEDWIKELQ CSKTIDAYAV ALLKFSELLP YALVADMTFE
wBm     HSSKRLLINN FDELFHLVNC SKEDHTKELQ RSKAIDECAI TLLKSSELLP YALVVDVNFK
        * *****  * ****   *  * ***   **** ** **  *   *** ***** **** *  *

        121                              > RibA  GTP cyclohydrolase II   180
wKue    NNHEMRNWCE KNDVIALDTS FINNFQENQD VYEVCKTSLF LKQTQEVNII SYRTESGGRE
wMel    NNHEMRNWCE KNDVIALDTS FINNFQENQD VYEVCKTSLF LKQTQEVNII SYRTESGGRE
wHa     NNHEMRNWCE KNDVIALDTS FINNFQENQD VYEVCKTSLF LKQTQEVNII SYRTESGGRE
wRi     NKYEMRNWCE ENDIIALDTL LVNDFQQNQS VYEVCKTSLF LKQTQEVDII SYRTESGGRE
wVulC   NNHEMQNWCE KNDVIALDTS FINNFQENQD VYEVCKTSLF LKQTQEVDII SYRTESGGRE
wStr    NNHEMRNWCE KNDVIALDKS FINNFQENQD VYEVCKTSLF LKQTQEVDII SYRTKSGGRE
wTai    NKHEMRNWCE ENDIIALNTL LVNDFQRNHS VYEVCKTSLF LKQTQEVDII SYRTKSGGRE
wPip    NKHEMRNWCE ENDIIALNTL LVNDFQQNHS VYEVCKTSLF LKQTQEVDII SYRTKSGGRE
wVitB   NKHEMRNWCE ENDIIALNTL LVNDFQQNHS VYEVCKTSLF LKQTQEVDII SYRTKSGGRE
wBm     DEYEMRGWCE KSDVIALDVL FINNFQQNQD IYEVCKTPLF LKQTQKVNII SYRTCNGRKE
        **   ***   * ***      * ** *  ****** ** ***** * ** **** *   *

        181> RibA  GTP cyclohydrolase II domain                      <240
wKue    HHAIIIGNPD KDDEPLVRIH SSCYTGDLLD SLSCDCRSQL HQAIQMIADS GSGIILYLMQ
wMel    HHAIIIGNPD KDDEPLVRIH SSCYTGDLLD SLSCDCRSQL HQAIQMIADS GSGIILYLMQ
wHa     HHAIIIGNPD KDDEPLVRIH SSCYTGDLLD SLSCDCRSQL HQAIQMIADF GSGIILYLMQ
wRi     HHAIIIGNPD KDDEPLVRIH SSCYTGDLLD SLSCDCRSQL HQAIQMIADF GSGIILYLMQ
wVulC   HHAIIIGNPD KDDEPLVRIH SSCYTGDLLD SLSCDCRSQL HQAIQMIADF GSGIILYLMQ
wStr    HYAIIIGNPD KDNEPLVRIH SSCYTGDLLD SLSCDCRSQS HQAIQIMTDF GNGIILYLMQ
wTai    HYAIIIGNPD KDNEPLVRIH SSCYTGDLLD SLSCDCRSQL HQAIQIMTDF GNGIILYLMQ
wPip    HYAIIIGNPD KDNEPLVRIH SACYTGDLLD SLSCDCRSQL HQAIQIMTDF GNGIILYLMQ
wVitB   HYAIIIGNPD KDNEPLVRIH SSCYTGDLLD SLSCDCRSQL HQAIQIMTDF GNGIILYLMQ
wBm     HYAIIIGNPG KNSEPLVRVH SSCYTGDLLD SLSCDCRSQL HQAIQIMTDS GNGIILYLMQ
        * *******  *  **** * * ******** ********* *****  *  ********

        241> RibA  GTP cyclohydrolase II domain                        300
wKue    DGRGIGLTNK LRAYSMQRGH NLDTVDANRI LGFEDDERSF AVAAKMLKKL NINKIQLLTN
wMel    DGRGIGLTNK LRAYSMQRGH NLDTVDANRI LGFEDDERSF AVAAKMLKKL NINKIQLLTN
wHa     DGRGIGLTNK LRAYSMQREH NLDTVDANRI LGFEDDERSF VVAAKMLKKL NINKIQLLTN
wRi     DGRGIGLTNK LRAYSVQREH NLDTVDANRI LGFEDDERSF VVAAKMLKKL NINKIQLLTN
wVulC   DGRGIGLANK LRAYSMQRRH NLDTVDANRV LGFEDDERSF AVAVEILKKL DIKKIQLLTN
wStr    DGRGIGLTNK LRAYSMQRKY NLDTVDANRV LGFEDDERSF AVAAKILKKL NINKIQLLTN
wTai    DGRGIGLTNK LRAYSMQRKY NLDTVDANRV LGFEDDERSF AVAAKILKKL NINKIQLLTN
wPip    DGRGIGLTNK LRAYSMQRKY NLDTVDANRV LGFEDDERSF AVAAKILKKL NINKIQLLTN
wVitB   DGRGIGLTNK LRAYSMQRKY NLDTVDANRV LGFEDDERSF AVAAKILKKL NINKIQLLKN
wBm     DGRGIGLTNK LRAYDMQRKY NLDTVDANRI LGFEDDERSF AVAAEMLKKL GIKKIQLLTN
        ********** ****  **   ******** ********* **   ****  * ***** *

        201> RibA  GTP cyclohydrolase II domain              <347
wKue    NDRKLSELES SGIGVTKCLP LIVERNKYND SYMETKFGKL GHRLRVF
wMel    NDRKLSELES SGIGVTKCLP LIVERNKYND SYMETKFGKL GHRLRVF
wHa     NDRKLSELES SGIEVTKCLP LIVERNKYND SYMETKFGKL GHKLRVF
wRi     NDRKLSELES SGIEVTKCLP LIVERNKYND SYMETKFGKL GHKLRVF
wVulC   NGRKLSELKN NGIEVTKCLP LIMERNEYND SYMETKFGRL GHGLRVF
wStr    NGRKLSELKN NGIEVTKCVP LIMERNEYND SYMETKFGKL GHGLRVY
wTai    NGRKLSELKN NGIEVTKCVP LIMERNEYND SYMETKFDKL GHGLRVY
wPip    NGRKLSELKN NGIEVTKCVP LIMERNEYNH SYMETKFGKL GHGLRVY
wVitB   NGRKLSELKN NGIEVTKCVP LIMERNEYND SYMETKFGKL DHGLRVY
wBm     NGRKLSELKN NGIEVTRCLP LIMERNKYND SYIETKFSRL GHRLRTF
        * ******   ** ** * * ** *** **  ** ****  *  * **
```
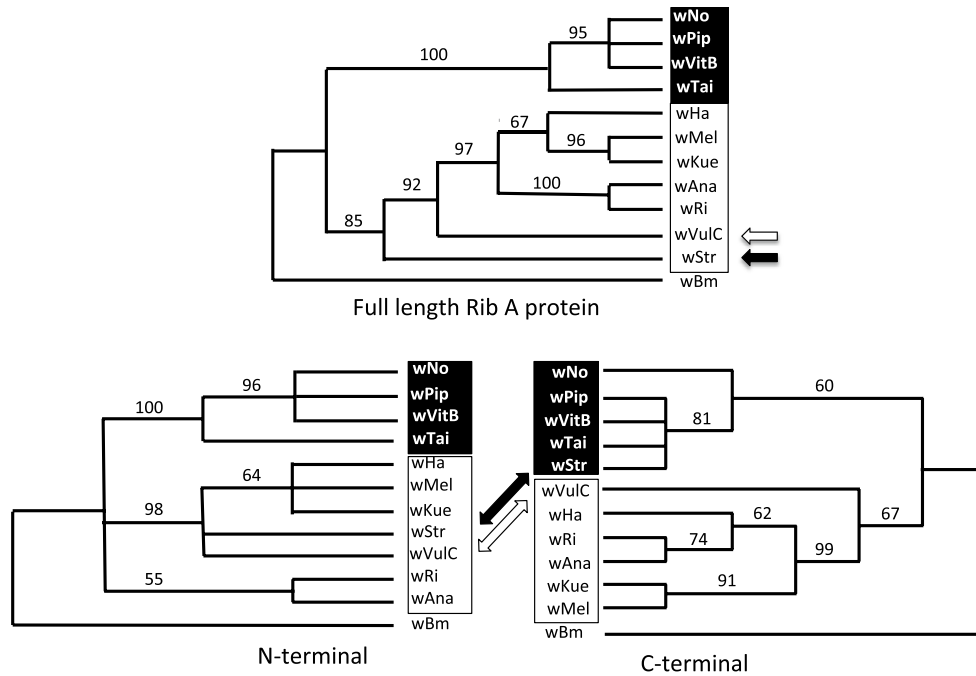
Fig. 4 Phylogenic relationships of [B]wStr RibA protein with homologs from WOL-A- and WOL-B-strains. Consensus trees show bootstrap values based on 1000 replicates, with [D]wBm (WOL-D) as the outgroup. WOL-A-strains are shown in *black font boxed* against a *white background*. WOL-B-strains are shown in *white font on a black background*. *Open arrows* designate [B]wVulC and *closed arrows* indicate [B]wStr. The N-terminal alignment corresponded to the first 150 residues in Fig. 3; the remainder of the protein was included in the C-terminal alignment

and [B]wVulC matched the other three A-strains. In aggregate, the alignment suggested that the [B]wStr and [B]wVulC homologs are two-part mosaics, each containing a protein functional domain, with an N-terminal WOL-A motif and a C-terminal WOL-B motif. We note that the C-terminal B-strain motif is consistent with the B-strain identity of the downstream *vir*B8-D4 operon (Table 2) and includes the predicted promoter region (Ohashi et al. 2002; Pichon et al. 2009). Likewise, in a phylogenetic comparison (Fig. 4), trees representing the full length and N-terminal regions (top and bottom left) show [B]wVulC and [B]wStr in adjacent positions, and grouped more closely with WOL-A-strains. In the C-terminus, where the amino acid alignment shows an overall higher consensus (Fig. 3), [B]wStr grouped with the B-strains including [B]wPip, while [B]wVulC appears more closely related to A-strains.

Nucleotide alignment and phylogenetic comparisons show that *rib*A is a mosaic gene in [B]wStr and [B]wVulC

A nucleotide alignment (Fig. S1) confirmed that *ribA* from [B]wStr is a two-part mosaic of WOL-A and WOL-B sequence motifs that correspond to the N- and C-terminal halves of the protein. In the first 522 nucleotides of *ribA*, 45 (in red font) of 56 variable nucleotides in [B]wStr match the A-strain sequences (Fig. S1), but only six (in blue) match

the majority of B-strains and two are unique to [B]wStr (in green). In the downstream 522 nucleotides of *rib*A, 51 (in blue) of 54 variable nucleotides in [B]wStr match B-strains, while a single nucleotide (684 in red) matches the A-strains and two (in green) are unique to [B]wStr. In [B]wVulC, *rib*A has a similar two-part mosaic structure but does not firmly transit from the WOL-A to the WOL-B sequence motif until position 775, consistent with the amino acid alignment. Among the A-strains, *rib*A from [A]wRi is again most similar to the B-strain sequences. Within nucleotides 387–453 encoding amino acids 129–150 just before the cyclohydrolase domain and the A/B-strain sequence motif transition in [B]wStr, 13 of 18 WOL-A/B variable nucleotides in [A]wRi are shared with [B]wTai, [B]wPip and [B]wVitB, but those of [B]wStr and [B]wVulC are conserved with the other A-strains (orange and black vs. red residues, respectively).

WspB in [B]wStr is strikingly similar to a [A]wCobU4-2 homolog

Having shown that *wsp*B is intact in [B]wStr, we mapped 11 peptides onto amino acid sequences encoded by 12 homologs (Fig. 5), including sequences deduced from three open reading frames (ORFs) in the *wsp*B pseudogene from [B]wPip (Sanogo et al. 2007) and two overlapping ORFs in a pseudogene from [A]wCobU4-2, one of several WOL-A

Fig. 5 Amino acid sequence alignment of WspB homologs. At *left*, font color designates WOL-A (*red*) and B (*blue*) strains, and the [B]*w*Str sequence is the top listed Wol-B-strain. *Asterisks below alignment indicate* universally conserved residues; three hypervariable regions (HVRs) are *doubly underlined above the alignment. Blocks of coloring* designate peptides detected by LC–MS/ MS at the 95 % confidence level. Those in *gray* were conserved in A- and B-strains. *Cyan* designates peptides conserved in B-strains, and *yellow*, those conserved in [B]*w*Str and [A]*w*CobU4-2. *Olive* peptides were unique to [B]*w*Str. Residues conserved between [B]*w*Str and a majority of A-strains are in *red font* (a single proline at residue 193) and residues conserved with a majority of B-strains are in *blue font*. Unique residues are in *green font*, and residues conserved between two or three homologs are in *orange font. Underlined residues below the alignment* denote the breakpoints between contiguous peptides within sequence regions. The *greater than* and *less than* symbols below the alignment indicate a transposon insertion in the *wsp*B pseudogene of [B]*w*Pip, followed by two additional deduced ORFs—see Fig. S2. PROFtmb (prediction of transmembrane beta barrels) symbols for individual residues below the alignment are: U— up-strand, D—down-strand, I—periplasmic loop, O—outer loop. PROFisis (prediction of protein–protein interaction residues) symbol P designates interaction residues. *Wolbachia* strain host associations: [A]*w*Atab 3, *A. tabida*—wasp; [A]*w*Cob, *C. obstrictus*—weevil; [B]*w*Met, *Metaseiulus occidentalis*—predatory mite. See Tables 2 and S2 for other host associations and GenBank Accessions. The first 20 residues of the [A]*w*Cob and [B]*w*Met sequences are not available

```
              1                                                   HVR1                   60
wAtab3    MISKKTLAVT AFALLLSQQS FASETEGFYF GSGYYGQYLN DTSVLKT--- --STTGIKNL
wKue      MISKKTLAVT AFALLLSQQS FASETEGFYF GSGYYGQYLN NTSVLKT--- --STTGIKNL
wMel      MISKKTLAVT AFALLLSQQS FASETEGFYF GSGYYGQYLN NTSVLKT--- --STTGIKNL
wRi       MISKKTLAVT AFALLLSQQS FASETEGFYF GSGYYGQYLN NTSVLKT--- --STTGIKNL
wAna      MISKKTLAVT ALALLLSQQS FASETEGFYF GSGYYGQYLN YTGELKAKIG DTAATATNNV
wCobU5-2  ---------- ---------- FASETEGFYF GSGYYGQYLN YMGELKAKIG DTAATAANNV
wCobU4-2  ---------- ---------- FASETEGFYF GGGYYGQYLN -LGKLKAKIG GKDATDDNHV
wStr      MMSKKTLAVT ALALLLSQQS FASETEGFYF GGGYYGQYLN -LGKLKAKIG GKDATDDNRV
wVitB     MMSKKTLAVT ALALLLSQQS FASETEGFYF GGGYYGQYLN -LGKLKAKIG GKDATDDNRV
wMet      ---------- ---------- FASETEGFYF GGGYYGQYLN -LGKLKAKIG GKDATDDNRV
wNo        MSKKTLAVT ALALLLSQQS FASETEGFYF GGGYYGQYLN -LGKLKAKIG DKDATDDNRV
wPip      -MSKKTLAVT ALALLLSQ-S FASETEGFYF GGGYYGQYLN -FGKLKAKIG SKDATDANKV
          * ******** ********** ********** *         *        **          *
PROFtmb   IIIIIIIIII IIIIIIIIII IIIIIIIUUU UUUUUUUOOO OOOOOOOOOO OOOOOOOOOO
PROFisis  PP-------- ---------- ---------- ---------- ---------- ----------

              61 HVR1          HVR1                                        120
wAtab3    SINDRGAQNT EGQSLSEYKG DYNPPFAANV AFGYTGELGN NSYRAELEGM YSSVKVDNIG
wKue      SINDRGAQNT EGQSLSEYKG DYNPPFAANV AFGYTGELGN NSYRAELEGM YSSVKVDNIG
wMel      SINDRGAQNT EGQSLSEYKG DYNPPFAANV AFGYTGELGN NSYRAELEGM YSSVKVDNIG
wRi       SINDRGAQNT EGQSLSEYKG DYNPPFAANV AFGYTGELGN NSYRAELEGM YSSVKVDNIG
wAna      SVNDRSAQNT EGQSLSKYKG DYNPPFAANV ALGYTGELNG NSYRAELEGM YSSVKVDNIG
wCobU5-2  SVNDRSAQNT EGQSLSKYKG DYNPPFAANV ALGYTGELNG NSYRAELEGM YSSVKVDNIG
wCobU4-2  SINDIDAQRT EGQLISKYKG DYNPPFAANV TFGYTGELGN NSYRAELEGM YSSVKVDNIG
wStr      SINDIDAQRT EGQLISKYKG DYKPPFAANV TFGYTGELGN NSYRAELEGM YSSVKVDNIG
wVitB     SINDIDAQRT EGQLISKYKG DYNPPFAANV TFGYTGELGN NSYRAELEGM YSSVKVDNIG
wMet      SINDIDAQRT EGQLISKYKG DYNPPFAANV TFGYTGELGN NSYRAELEGM YSSVKVDNIG
wNo       FINDRNTERT EPQPISEYKA DYSPPFAANI AFGYTGELGN NSYRAELEGI YSSIKVNNIG
wPip      SVNDRGAQST EGQLLNKYEG DYNPPFAANV ALAYTGELGN NSYRAELEGM YSSVKVDNIR
          **        * * *      *  ** *****   ******* ********* *** *****
PROFtmb   OOOOOOOOOO OOOOOOOOOO OOOOOODDDD DDDDDDDDDI IIUUUUUUUU UUUUOOOOOO
PROFisis  ---------- ---------- ---------- -------P-- ---------- ----------

              121 HVR2                     HVR2                            180
wAtab3    LTSSQITVSY LKETGEDPNK ETYLYSAAVS HDQIENISVM ANVYHHWKSD RFSFSPYVGI
wKue      LTSSQITVSY LKETGEDPNK ETYLYSAAVS HDQIENISVM ANVYHHWKSD RFSFSPYVGI
wMel      LTSSQITVSY LKETGEDPDK ETYLYSAAVS HDQIENISVM ANVYHHWKSD RFSFSPYVGI
wRi       LTSSQITVSY LKETGEDPNK ETYLYSAAVS HDQIENISVM ANVYHHWKSD RFSFSPYVGI
wAna      LTSGQMTISY TKDK-TRPEE -SY--GAIVN HDQIENISVM ANVYHHWKSD RFSFSPYVGV
wCobU5-2  LASSQITISY IKDK-ANPEE -RY--GAIVN HDQIENASLM ANVYHHWKSD RFSFSPYVGI
wCobU4-2  LSSNQVTVSY LKDVGESANK KTYMYKTVIN HDQVENASVM ANVYHYWKSD SFSFSPYVGI
wStr      LSSNQVTVSY LKDVGESANK KTYMYKTVIN HDQVENASVM ANVYHYWKSD SLSFSPYVG
wVitB     LSSNQVTVSY LKDVGESTNK KTYMYKTVIN HDQVENASVM ANVYHYWKSD SFSFSPYVGI
wMet      LSSNQVTVSY LKDVGESTNK KTYMYKTVIN HDQVENASVM ANVYHYWKSD SFSFSPYVGI
wNo       LANTQMNIKY -------EK ENNKYGVTIN HGKIDNISVM ANVYHHWKND SFSFSPYVGI
wPip      LTSGQMTMDQ TEGGNQTMDQ FLTMTKYL-- -------SVM ANVYHHWKSE SFSFSPYVGI
          * * *      *                >  *** *<* * ***** *** *******
PROFtmb   OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOODDDD DDDDDDDDDI IUUUUUUUUU
PROFisis  ---------- --------PP PP-------- ---------- -------PP P---------

              181                                         HVR3         240
wAtab3    GIGATRMTMF EKPSIRPAGQ LKAGFDYRIN EDVNMHIGYR GFGAIGSDIK LTAKRLGQVV
wKue      GIGATRMTMF EKPSIRPAGQ LKAGFDYRIN EDVNMHIGYR GFGAIGSDIK LTAKRLGQVV
wMel      GIGATRMTMF EKPSIRPAGQ LKAGFDYRIN EDVNMHIGYR GFGAIGSDIK LTAKRLGQVV
wRi       GIGATRMTMF EKPSIRPAGQ SKAGFDYRIN EDVNMHIGYR GFGAIGSDIK LTAKRLGQVV
wAna      GVGATRMTMF EKSSIRPAGQ LKAGLDYRIN EDVNMHIGYR GFGAIGS--- -SEYKLDTLK
wCobU5-2  GVGATRMKMF EKSSIRPAGQ LKAGFDYRIN EDVNRHIGYR GFGVLGSNVD FEAEVLGEMK
wCobU4-2  GVGATRMTMF EKSSIRPAGQ LKAGFDYRIN EDVNRHIGCR GFGVLGSNVD FEAEVLGEMK
wStr      GVGATRMTMF EKPSIRPAGQ LKAGFDYRIN EDVNMHIGYR GFGVLGSNVD FEAEVLGEMK
wVitB     GVGGTRMTMF EKSSIRPAGQ LKAGFDYRIN EDVNMHIGYR GFGVLGSNVD FEAEVLGEMK
wMet      GVGGTRMTMF EKSSIRPAGQ LKAGFDYRIN EDVNMHIGYR GFGVLGSNVD FEAEVLGEMK
wNo       GVGATRMTMF EESSIRPAGQ LKAGFDYHIN EDVNMHIGYR GFGVIGS--- -SEYKPETLK
wPip      GVGATRMTMF EKSSIRPAGQ LKAGFDYRIN EDVNMHIGYR GFGVLG---- ----------
          * * *** ** *   ****** ******* ** **** ***** ***   **
PROFtmb   UUUUUUOOOO OOOOOOODDD DDDDDDDDDD IIUUUUUUUU UUUUOOOOOO OOOOOODDDD
PROFisis  ---------- P-P------- ---------- ---------- ---------- ----------

              241 HVR3                     HVR3                     295
wAtab3    DDPNNDKKK- -------KLN PSSGSKVTEE INIGNQLFHT HGIEAGLTFH FASKA
wKue      DDPNNDKKK- -------KLN PSSGSKVTEE INIGNQLFHT HGIEAGLTFH FASKA
wMel      DDPNNDKKK- -------KLN PSSGSKVTEE INIGNQLFHT HGIEAGLTFH FASKA
wRi       DDPNNDKKK- -------KLN PSSGSKVTEE INIGNQLFHT HGIEAGLTFH FASKA
wAna      WDPNHDNGKD KPKGGMAEQT ----GDNQVS TTIQNDFFHT HGIEAGLTFH FASKA
wCobU5-2  VKQQVNPDGK KILELNKSQK PSDQKLHKES ISIGNQVFHT HGIEAGLTFH FASKA
wCobU4-2  VKQQVNPDGK KILELNKSQK PSDQKLHKES ISIGNQVFHT HGIEAGLTFH FASKA
wStr      AKQPVNPDGK KILELNKSQK PSDQKLHKES ISIGNQVFHT HVIEAGLTFH FASKA
wVitB     AKQQVNQDGK KILELNKNQK PSDQKLYKES ISIGNQVFHT HGIEAGLTFH FASKA
wMet      EKQQANSDGK KILELNKNQK PSDQKLHKES ISIGNQVFHT HGIEAGLTFH FASK-
wNo       -------LNA KTKKMNKQIG ---ENKVT-- AAIQNSFFHT HGIEAGLTFH FASKA
wPip      ---------- ---------- ---------- -------FHT HGIEAGLTFH FASKS
          **                            *   *** ********** ****
PROFtmb   DDDDDDDDII UUUUUUUUUU UUOOOOOOOO OOOOOOOOOO DDDDDDDDDD DDIII
PROFisis  PP-------- ------P-PP -PPP------ ---------- ---------- -----
```

variants associated with the weevil, *Ceutorhynchus obstrictus*. Of two [B]*w*Str peptides (Fig. 5) detected at 95 % confidence in the original search (Baldridge et al. 2014), the first (residues 105–115 in gray) was identical in all strains except [B]*w*No, which has unique M/I and V/I substitutions (residues in green). The second peptide (residues 209–220) is identical in all but the two [A]*w*Cob strains that share an M/R substitution (215 in orange), while [A]*w*CobU4-2 has a unique Y/C substitution (219 in green). Five additional [B]*w*Str peptides (highlighted in cyan) were identical with [B]*w*VitB and [B]*w*Met (residues in blue), but not with [B]*w*Pip and [B]*w*No, which have many residues that are unique (in green) or shared (in orange) only with [A]*w*CobU5-2 and [A]*w*Ana. Thus, with the exception of [A]*w*CobU5-2, cyan peptides of [B]*w*Str match other WOL-B-strains.

Two peptides underscore a striking similarity between the [B]*w*Str and [A]*w*CobU4-2 homologs. The first (Fig. 5, residues 133–140 highlighted in yellow) contains an alanine residue (138 in bold orange) shared only with [A]*w*CobU4-2. The second (residues 169–186 highlighted in olive) has a unique F/L substitution (in green) and a V/I substitution (in orange) shared with [A]*w*CobU4-2 and [A]*w*Ana. Overall, the [B]*w*Str and [A]*w*CobU4-2 sequences differ at only five residues (59, 172, 193, 215 and 219), of which four occur within hypervariable regions. Throughout the alignment, [A]*w*Atab 3, [A]*w*Kue, [A]*w*Mel and [A]*w*Ri form a conserved group, but the divergent [A]*w*Ana and [A]*w*CobU4-2 and U5-2 strains have multiple residues (in blue, as in 42–77 and 224–277) that are conserved with the B-strains, suggesting genetic exchange between supergroups.

WspB domain structure and hypervariable regions (HVRs)

WspB is a paralog of the better-known WspA major surface antigen, which is anchored in the cell envelope by a transmembrane β-barrel domain (Koebnik et al. 2000), while surface-exposed loop domains contain HVRs with high recombination frequencies within and between strains (Baldo et al. 2010). The PROFtmb program predicted 10 transmembrane down (D)- and up (U)-strands and six periplasmic space (I) strands in WspB from [B]*w*Str (Fig. 5; residues indicated by D, U and I, respectively; Z score of 6.8 supports designation as transmembrane β-barrel protein). HVR1 and HVR2 each contain a predicted outer loop (residues 38–86 and 115–156 indicated by O) with high proportions of amino acids that are potentially charged at physiological pH; HVR3 contains two outer loops. Finally, a small predicted loop that is not within an HVR contains a proline (residue 193) that is conserved in [B]*w*Str and four WOL-A-strains. It is one of the 20 amino acids, most with hydrophilic or potentially charged side chains and within HVRs or adjacent to periplasmic space strands, predicted

by the PROFisis program to be potentially involved in protein–protein interactions (P below alignment).

HVR1 amino acids

In HVR1 (Fig. 5, residues 41–77), eight residues are universally conserved among all homologs, while the majority of variable residues are differentially conserved in the B-strains (residues in blue) versus the A-strains. However, the sequences from the [A]*w*Ana and [A]*w*CobU5-2 A-strains are mosaics in which eight of the first 20 residues (in blue) are conserved with all B-strains, while eight others are either conserved mutually or with [B]*w*No or [B]*w*Pip (in orange). Within the remaining 17 residues of HVR1, the [A]*w*Ana and [A]*w*CobU5-2 sequences are better conserved with the other A-strains, while [B]*w*No and [B]*w*Pip have multiple unique residues (in green). The [A]*w*CobU4-2 and [B]*w*Str sequences differ only at residue 59.

HVR2 amino acids

Within HVR2 (Fig. 5, residues 121–150), [A]*w*CobU5-2 and [A]*w*Ana sequences have alignment gaps at four residues, five or six unique residues respectively (in green), and eight residues that are either conserved mutually (in orange) or with [B]*w*No. The [B]*w*Pip pseudogene has only the first two residues of HVR2 due to a transposon insertion (indicated below alignment by *greater than less than* symbols). The [A]*w*CobU4-2 pseudogene contains a nucleotide sequence duplication (see below) that results in an overlap of the first and third ORFs beginning at the seventh residue of HVR2, but their spliced sequences, as shown, are identical to that of [B]*w*Str. The [B]*w*No sequence has eight alignment gaps and nine unique residues.

HVR3 amino acids

In HVR3, five of 52 residues (Fig. 5, residues 224–277) are conserved among all strains. Throughout HVR3, sequences from the upper cluster of four A-strains are identical, including an alignment gap. However, the [A]*w*Ana sequence has 22 unique residues (in green) and is partially conserved with [B]*w*No (nine residues in orange). In striking contrast to differences in HVR1 and HVR2, the [A]*w*CobU4-2 and U5-2 homologs have identical HVR3 sequences that are conserved with the B-strains, particularly [B]*w*Str (residues in blue), differing only at residues 241 and 244.

Nucleotide sequence alignment confirms a mosaic *wsp*B and identifies a conserved repeated sequence

Nucleotide sequence alignment of eleven *wsp*B homologs confirmed that WOL-A/B genetic mosaicism is

concentrated in the HVR regions and revealed three copies of a repeated sequence element within or near HVR2. Further analyses identified three copies of the repeated sequence element in *rib*A at the 5′-end of the *vir*B8-D4 operon and four copies in *vir* genes.

HVR1

HVR1 (Fig. S2, nucleotides 117–241) from [B]*w*Str begins with two nucleotides (117 and 120 in red) that are conserved in [B]*w*Str and all WOL-A-strains except [A]*w*CobU5-2 and [A]*w*CobU4-2. Downstream, the [B]*w*Str sequence includes 47 of 48 nucleotides (in blue) within a sequence motif characteristic of [B]*w*Str and the other B-strains. The [A]*w*CobU5-2 and [A]*w*Ana sequences are initially similar to the WOL-B motif, but beginning at an alignment gap in the other A-strains they have 11 nucleotides (in orange, nucleotides 152–207) that are conserved with [B]*w*No and [B]*w*Pip at positions in which those strains diverge from the WOL-B consensus. Thus, HVR1 in [B]*w*Str begins with nucleotides from a conserved WOL-A sequence motif but transitions to the conserved WOL-B motif, while HVR1 from the [A]*w*CobU4-2 A-strain differs from that WOL-B motif at a single nucleotide (176). In contrast, the [A]*w*Ana and [A]*w*CobU5-2 sequences are mosaics of the WOL-A and WOL-B consensus motifs and share nucleotides with the divergent [B]*w*No and [B]*w*Pip B-strains, which also closely resemble each other upstream of HVR1 (23 nucleotides in light blue and one in orange).

HVR2 contains conserved repeat elements

HVR2 (Fig. S2, nucleotides 361–450) contains a conserved WOL-B sequence motif that differs at 20 nucleotides (in blue), from the WOL-A motif, while the divergent sequences from [B]*w*No, [B]*w*Pip, [A]*w*Ana and [A]*w*CobU5-2 share an alignment gap and are again similar (nucleotides in orange). A tandem repeated sequence at nucleotides 365–379, **CAAGTAATCAAGTAAC**, in the B-strains [B]*w*Str, [B]*w*VitB and [B]*w*Met occurs with slight variation (underlined residues) as **CAAGTAGCCAAATAAC**, in the A-strains [A]*w*Atab 3, [A]*w*Kue, [A]*w*Mel and [A]*w*Ri. We designated the eight-bp sequence, CAARTARY, where R = A or G, and Y = C or T, as an HVR2-repeat. The pseudogene from [A]*w*CobU4-2 contained a third copy of CAAGTAAT that interrupted ORF1 and was removed from the alignment (indicated by upwards arrow below alignment) to shift to ORF3, which maintains identity to the deduced amino acid sequence from [B]*w*Str. Just downstream of HVR2 at nucleotides 457–463, a truncated copy of the HVR2-repeat lacking the 3′-terminal pyrimidine is conserved in [B]*w*Str, [B]*w*VitB, [B]*w*Met and [A]*w*CobU4-2 and corresponds to the position (indicated by *greater than less than* symbols

below alignment) of the transposon insertion in [B]*w*Pip. Finally, we noted that the most divergent HVR2 sequences from [A]*w*Ana, [A]*w*CobU5-2, [B]*w*No and [B]*w*Pip have T/C and A/G substitutions (in orange, light blue and green) that disrupt the HVR2-repeat consensus.

HVR3

Within HVR3 (Fig. S2, nucleotides 670–831), conserved sequence motifs occur in the upper cluster of four A-strains and in the B-strains (nucleotides in blue), with the exceptions of [B]*w*Pip (HVR3 absent) and [B]*w*No. Sequences from [A]*w*CobU4-2 and [A]*w*CobU5-2 are identical despite their major differences in HVR1 and HVR2 and differ from the B-strain consensus only at nucleotides 722 and 773 (in orange). The [A]*w*Ana and [B]*w*No sequences are the most divergent but share 43 variable nucleotides (in orange) and have 67 and 18 unique residues (in green), respectively.

HVR2-repeats also occur in *rib*A and *rib*B

Based on a DNA pattern search (http://bioinformatics.org/sms/), three HVR2-repeats occur in *rib*A, two in *vir*D4, and single copies in *vir*B8 and *vir*B9 (Table 3). In addition, a reverse complement of the CAARTARY sequence occurs at the same position in *rib*B from three WOL-A-strains and [B]*w*Pip (see gray shading in Fig. S3). The [B]*w*Pip homolog contains a second copy at residues 7–14 just downstream of the start codon (not shown) and is a WOL-A/B mosaic (see below). Although repeat frequencies in individual *rib*A (0.29) and *wsp*B (0.34) genes are ~sixfold higher than in the whole genomes of [A]*w*Mel and [B]*w*Pip (0.05) from flies (Diptera), it will be important to re-evaluate these frequencies when a [B]*w*Str genome (Hemipteran host) becomes available.

Although RibA and RibB are involved in riboflavin biosynthesis, *rib*B is not contiguous with *rib*A and the *vir*V8-D4 operon, and it has higher variability than *rib*A (Table 2). Among the WOL-B-strains, *rib*B in [B]*w*Str and [B]*w*No is conserved with the [A]*w*Au and [A]*w*Mel A-strains (Fig. S3; note especially the bold blue residues downstream of nucleotide 181, as well as additional residues in orange). In contrast, the [B]*w*Pip homolog is best-conserved (nucleotides in red) with WOL-A-strains, [A]*w*Ana, [A]*w*Ha and [A]*w*Ri, including an alignment gap at residue 483 encompassing an identical 15-nucleotide "island" with the reverse complement CAARTARY repeat. Downstream of the gap, at residue 511, the [B]*w*Pip sequence shifts to a predominantly WOL-B motif conserved in [B]*w*Str, [B]*w*No, but also in [A]*w*Mel (nucleotides in blue), while [A]*w*Ana, [A]*w*Ri and [A]*w*Ha are mutually conserved (nucleotides in orange) versus all other strains. Within the 3′-end of the alignment (nucleotides 541–600), the [B]*w*Pip sequence is conserved with [B]*w*Str, [B]*w*No and

Table 3 Distribution of HVR2-repeats in [B]wStr virB8-D4 operon and genomes of [A]wMel and [B]wPip

Repeat	ribA	virB8	virB9	virD4	wspB	wMel	wPip
CAAGTAAT/C	118/145	–	–	−5943[b]	7610/7618	154	239
CAAATAAT/C	672	–	−2485	−5919[b]	–	275	360
CAAGTAGC	–	1288	–	–	7702[b]	187	104
Total	3	1	1	2	3	616	703
Frequency[a]	0.29	0.15	0.13	0.12	0.34	0.05	0.05

Values indicate 5′-nucleotide positions of HRV2-repeats in the 9133-bp ribA to topA sequence from [B]wStr (see Fig. 1; Acc. KF43064.1). Negative values indicate reverse complement positions. Copy numbers in the complete [A]wMel (NC_002978.6) and [B]wPip (NC_010981.1) genomes are shown at right

[a] Frequency is defined as number repeats/total nucleotides in each individual gene (or complete genome) indicated at the top of the panel, ×100

[b] See underlined nucleotides 457–463 in Fig. S2, which lack the 3′-terminal pyrimidine

[D]wBm (nucleotides in blue), while [A]wAu and [A]wMel are the most divergent (nucleotides in green).

Discussion

Although the status of *Wolbachia* as a species remains unclear (Baldo et al. 2006b; Lo et al. 2007), a notable distinction between WOL-C-/D-strains that associate with nematodes as mutualists and WOL-A-/B-strains that occur as reproductive parasites in insects relates to genome stability and phylogenetic congruence between *Wolbachia* and its host. In insect hosts, *Wolbachia* appears to engage in frequent horizontal gene transfer, resulting in a lack of phylogenetic congruence manifested by gene structures that represent mosaic recombinations from genomes now considered distinct strains. Coinfections with two or more *Wolbachia* strains and activities of bacteriophages that reside in genomes of WOL-A/B-strains likely contribute to this genetic plasticity (Bordenstein and Reznikoff 2005; Newton and Bordenstein 2011), which may reflect what some authors suggest is a worldwide *Wolbachia* pandemic (Zug et al. 2012). Examples of natural coinfections include [A]wAlbA and [B]wAlbB in *A. albopictus* mosquitoes (O'Neill et al. 1997), [A]wVitA and [B]wVitB in the parasitoid wasp, *N. vitripennis* (Perrot-Minnot et al. 1996; Raychoudhury et al. 2008) and [A]wHa and [B]wNo in the phytophagous *D. simulans* (James et al. 2002). A particularly interesting example in *C. obstrictus* weevils involves infection with a single [A]wCob strain, in which polymorphisms in *wsp*A and *wsp*B indicate that three distinct variants coexist in all host populations (Floate et al. 2011) and it will be of interest to explore other genetic similarities and differences among these variants following separation in vitro and/ or in uninfected hosts. *Wolbachia* coinfections have also been documented in insects such as fig wasps (Yang et al. 2012), tephritid flies (Morrow et al. 2014) and planthoppers (Zhang et al. 2013) whose interactions with parasitoids,

parasites and predator arthropods may facilitate horizontal transmission (Cordaux et al. 2001; Werren et al. 2008; Zug et al. 2012). In nature, the [B]wStr strain occurs in two planthopper hosts (Noda et al. 2001a) and in the strepsipteran endoparasite *Elenchus japonicus* (Noda et al. 2001b; Zhang et al. 2013). In the present study, [B]wStr has been artificially introduced into a cultured cell line, which has not been achieved with [B]wPip or nematode-associated strains. Adaptation of [B]wStr to cell lines (Noda et al. 2002; Fallon et al 2013) will provide an in vitro system for examining mechanisms of genetic exchange if conditions for maintenance of doubly infected cells can be developed through coinfection or somatic cell fusion. We note that high rates of recombination and transposition in *Wolbachia* (Baldo et al. 2006a; Cordaux et al. 2008) are consistent with expression of an abundant RecA protein (SR 1.05; Table S3, entry 146) as well as 18 transposases and/or proteins with transposase domains in [B]wStr (Baldridge et al. 2014).

Genetic plasticity of *wsp*B in the *vir*B8-D4 operon

An intact *wsp*B that maps to the 3′-end of the *vir*B8-D4 operon in most WOL-A genomes (Wu et al. 2004) is absent from 17 of 21 WOL-B-strains, including [B]wVulC and nearly all other isopod-associated strains (Pichon et al. 2009), and is interrupted by a transposon in [B]wPip (Sanogo et al. 2007). Here, we verify that in [B]wStr, an intact *wsp*B is co-transcribed with *vir*D4 and is expressed in C/wStr1 cells as an abundant protein at levels similar to those of many housekeeping proteins. The *wsp*B structure closely resembles that of its better-studied *wsp*A paralog, encoding a major surface antigen that has four HVR regions with sequence motifs that have been shuffled by recombination within and between *Wolbachia* WOL-A- and -B-strains (Baldo et al. 2005, 2010). Likewise, most sequence variation in *wsp*B alleles occurs in the three HVR regions, with distinctive patterns for each region. HVR1 underscores WOL-A/B mosaicism in [A]wAna and [A]wCobU5-2,

and in addition it shows a high level of identity between [A]wCobU4-2 and [B]wStr. Similarity between [A]wAna and [A]wCobU5-2 and between [B]wStr and [A]wCobU4-2 also occurs in HRV2, while [B]wNo stands out as distinctive. In [B]wPip, HVR2 is disrupted by a transposon insertion and we identified an eight-nucleotide HRV2-repeat (CAARTARY) that correlates with transitions between WOL-A-/B-strain motifs and the pseudogene lesions in [B]wPip and [A]wCobU4-2. Finally, we noted that high identity of [A]wCobU5-2, [A]wCobU4-2 and [B]wStr is unique to HVR3.

The remarkable similarity of the *wsp*B homologs from [B]wStr and [A]wCobU4-2 (>98 % nucleotide identity Fig. S2) is consistent with exchange of an apparently intact gene between members of distinct *Wolbachia* supergroups by a mechanism that requires further investigation. Intensive analysis of the *wsp*A paralog demonstrates that intragenic recombination breakpoints are concentrated in conserved regions outside of the HVRs (Baldo et al. 2005, 2010). CAARTARY repeats are not present in *wsp*A, and in *wsp*B, they occur only within and directly adjacent to HVR2 at positions that correspond to pseudogene lesions in [A]wCobU4-2 and in [B]wPip (due to a transposition event in [B]wPip; Sanogo et al. 2007). Furthermore, Pichon et al. (2009) suggested that transposition events may explain absence of *wsp*B in the *vir*B8-D4 operons of many WOL-B-strains. In a practical sense, CAARTARY repeats at *wsp*B pseudogene lesions and WOL-A/B sequence motif transitions (Figs. S1, S2, S3) suggest their involvement in genetic exchange. Because transformation of *Wolbachia* has not yet been achieved, engineering of CAARTARY repeats into vectors used successfully to introduce selectable markers into other members of the *Rickettsiales* (see Beare et al. 2011) merits investigation.

Potential functions of WspB

Although bacterial outer membrane proteins are important mediators of interactions with host cells and specific function(s) of both WspA and WspB remain to be identified, they may have unique functions as porin proteins in *Wolbachia*, which lack cell walls. The *vir*B8-D4 operons of *Wolbachia* and its sister genera, *Anaplasma* and *Ehrlichia*, are similarly organized (Gillespie et al. 2010; Hotopp et al. 2006) with 3′- terminal genes encoding major surface proteins that, analogous to *wsp*B, are co-transcribed with the *vir* genes (Ohashi et al. 2002). In *A. marginale*, a family of *msp*2 pseudogenes undergo "combinatorial gene conversion" at the expression site (Brayton et al. 2002) and MSP2 variants change during growth in different host cell types, which likely reflects a response to host immunity mechanisms (Chávez et al. 2012). Similarly, Baldo et al. (2010) proposed that changes in WspA HVR regions play a role in host adaptation and innate immunity interactions,

consistent with variation in the higher-order structure of the protein in different hosts (Uday and Puttaraju 2012). HVR sequence changes in the *wsp*B paralog may reflect a similar dynamic. Additional evidence indicates that MSP2 proteins are glycosylated (Sarkar et al. 2008), which is now an established process in post-translational modification in bacteria (Dell et al. 2010; Nothaft and Szymanski 2010), and we note that WspB contains potential glycosylation sites. Although an inactivated pseudogene or absence of *wsp*B in *vir*B8-D4 operons of some *Wolbachia* strains indicates that it is not absolutely required for survival, a secretome analysis of *Brugia malayi* showed that WspB from [D]wBm is excreted/secreted into filarial host cells (Bennuru et al. 2009). Furthermore, it co-localizes with the Bm1_46455 host protein in tissues that include embryonic nuclei (Melnikow et al. 2011). WspB is therefore itself a candidate T4SS effector that may play a role in reproductive manipulation of the host. Mosaicism in *wsp*B and its high rate of evolution (Comandatore et al. 2013) may thus reflect genetic changes that optimize adaptation to particular host cells such as those in reproductive tissues and facilitate exploitation of new arthropod niches by *Wolbachia*.

Genetic plasticity of *rib*A in the *vir*B8-D4 operon

Aside from *wsp*B at the 3′-end of the T4SS *vir*B8-D4 operon, *rib*A exhibits genetic plasticity at its 5′-end. In both [B]wStr and [B]wVulC, *rib*A is a two-part mosaic of N-terminal WOL-A and C-terminal WOL-B motifs. In contrast, the internal *vir*B8-D4 genes have typical B-strain identities, and in some strain comparisons, amino acid identities slightly exceed nucleotide identities, which Pichon et al. (2009) attribute to strong selection against non-synonymous codon substitutions. Among the internal *vir*B8-D4 genes, however, Klasson et al. (2009) suggest that in [A]wRi, an especially variable region in *vir*B10 is likely derived from genetic exchange with a B-strain. We note here that *rib*A from [A]wRi closely resembles B-strain homologs within a variable region that immediately precedes the GTP cyclohydrolase domain, where its homolog in [B]wStr transitions from WOL-A to WOL-B sequence motifs (Fig. S1, positions 387–450).

In contrast to [D]wBm, in which *rib*A and *vir*B8 are co-transcribed and bind common transcription factors (Li and Carlow 2012), relative abundance levels suggest that in [B]wStr, *rib*A is transcribed independently of the *vir*B8-D4 operon. Some WOL-B-strains, such as [B]wVulC, lack *wsp*B at the 3′-terminus of the *vir*B8-D4 operon, while our data confirm that in [B]wStr, *wsp*B is co-transcribed with the *vir* genes, consistent with similar relative abundances of WspB and the five Vir proteins. In aggregate, these observations suggest that WOL-D and WOL-A-/B-strains may differ in how RibA and WspB expression interfaces with

T4SS-mediated transport of effectors in filarial worms and arthropod hosts (Felix et al. 2008; Masui et al. 2000; Rances et al. 2008; Wu et al. 2004), and it will be of interest to explore whether such differences relate to riboflavin provisioning. In filarial nematodes (Li and Carlow 2012; Strubing et al. 2010; Wu et al. 2009) and bedbugs (Hosokawa et al. 2010), evidence suggests that *Wolbachia* provisions host with riboflavin, the precursor of flavin cofactors that are essential for many cellular redox reactions. In contrast, riboflavin depletion reduces [B]*w*Str abundance in C/*w*Str1 cells, suggesting that [B]*w*Str utilizes host riboflavin and does not augment riboflavin levels in mosquito host cells (Fallon et al. 2014).

Potential functions of RibA and RibB

In initial commitment steps in riboflavin biosynthesis, enzymatic activities encoded by the *rib*A and *rib*B functional domains use GTP and ribulose-5-phosphate as substrates to catalyze riboflavin biosynthesis, consuming 25 molecules of ATP per molecule of riboflavin (Bacher et al. 2000). We note that in *Wolbachia* genomes, *rib*A is the annotated homolog of *rib*BA in *Escherichia coli* (Brutinel et al. 2013) and encodes a dihydroxybutanone phosphate synthase domain with putative RibB function near the N-terminus, upstream of a GTP cyclohydrolase II domain with conserved dimerization and active site residues (RibA function). As in *E. coli*, *Wolbachia* genomes also encode *rib*B, but at a distinct chromosomal locus, suggesting that *rib*A and *rib*B are not coordinately expressed. In *Sinorhizobium meliloti* (*Rhizobiales*; *Alphaproteobacteria*), knockout mutations of *rib*BA decreased flavin secretion but did not cause riboflavin auxotrophy or block establishment of symbiosis, suggesting that RibBA may have an undefined role in molecular transport (Yurgel et al. 2014). As is the case with [B]*w*Str, RibB is at least threefold more abundant than RibA in the bacterium *Acidithiobacillus ferrooxidans* (Knegt et al. 2008). In yeast, RibB has thiol-dependent alternative redox states (McDonagh et al. 2011), partially localizes to the mitochondrial periplasm, and has an unexplained function in oxidative respiration that is independent of riboflavin biosynthesis (Jin et al. 2003). These observations raise the possibility that in *Wolbachia*, RibA and RibB may have functions other than riboflavin biosynthesis that integrate with pathways involved in cellular oxidative state, such as iron metabolism. Intracellular bacteria are challenged by host-imposed oxidative stress and iron starvation (reviewed by Benjamin et al. 2010) and riboflavin biosynthesis is associated with iron acquisition in bacteria such as *Helicobacter pylori* (Worst et al. 1998) and *Campylobacter jejuni* (Crossley et al. 2007). *Wolbachia* interferes with iron metabolism and sequestration in insects (Brownlie et al. 2009; Kremer et al. 2009) and influences iron-dependent

host processes such as heme metabolism, oxidative stress, apoptosis and autophagy (Gill et al. 2014). We note that the periplasmic iron-binding component of a membrane transporter is an abundant protein in [B]*w*Str (Table S3, entry 778 and Baldridge et al. 2014).

Compliance with ethical standards

References

Alvarez-Martinez CE, Christie PJ (2009) Biological diversity of prokaryotic type IV secretion systems. Microbiol Mol Biol Rev 73:775–808

Bacher A, Eberhardt S, Fischer M, Kis K, Richter G (2000) Biosynthesis of vitamin B2 (riboflavin). Annu Rev Nutr 20:153–167

Baldo L, Lo N, Werren JH (2005) Mosaic nature of the *Wolbachia* surface protein. J Bacteriol 187:5406–5418. doi:10.1128/JB.187.15.5406-5418.2005

Baldo L, Bordenstein S, Werengreen JJ, Werren JH (2006a) Widespread recombination throughout *Wolbachia* genomes. Mol Biol Evol 23:437–449

Baldo L, Dunning Hotopp JC, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, Hayashi C, Maiden MC, Tettelin H, Werren JH (2006b) Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. Appl Environ Microbiol 72:7098–7110

Baldo L, Desjardins CA, Russell JA, Stahlhut JK, Werren JH (2010) Accelerated microevolution in an outer membrane protein (OMP) of the intracellular bacteria *Wolbachia*. BMC Evol Biol. doi:10.1186/1471-2148-10-48

Baldridge GD, Baldridge AS, Witthuhn BA, Higgins L, Markowski TW, Fallon AM (2014) Proteomic profiling of a robust *Wolbachia* infection in an *Aedes albopictus* mosquito cell line. Mol Microbiol 94:537–556. doi:10.1111/mmi.12768

Beare PA, Sandoz KM, Omsland A, Rockey DD, Heinzen RA (2011) Advances in genetic manipulation of obligate intracellular bacteria. Front Microbiol 2:97. doi:10.3389/fmicb.2011.00097

Benjamin JA, Desnoyers G, Morissette A, Salvail H, Massé E (2010) Dealing with oxidative stress and iron starvation in microorganisms: an overview. Can J Physiol Pharmacol 88:264–272

Bennuru S, Semnani R, Meng Z, Ribeiro JMC, Veenstra TD, Nutman TB (2009) *Brugia malayi* excreted/secreted proteins at the host/parasite interface: stage-and gender-specific proteomic profiling. PLoS Negl Trop Dis 3:e410. doi:10.1371/journal.pntd.0000410

Bigelow HR, Petrey DS, Liu J, Przybylski D, Rost B (2004) Predicting transmembrane beta-barrels in proteomes. Nucleic Acids Res 32:2566–2577

Bordenstein SR, Reznikoff WS (2005) Mobile DNA in obligate intracellular bacteria. Nat Rev Microbiol 3:688–699

Bourtzis K (2008) *Wolbachia*-based technologies for insect pest population control. Adv Exp Med Biol 627:104–113. doi:10.1007/978-0-387-78225-6_9

Brayton KA, Palmer GH, Lundgren A, Yi J, Barbet AF (2002) Antigenic variation of *Anaplasma marginale msp*2 occurs by combinatorial gene conversion. Mol Microbiol 43:1151–1159

Brownlie JC, Cass BN, Riegler M, Witsenburg JJ, Iturbe-Ormaetxe I, McGraw EA, O'Neill SL (2009) Evidence for metabolic provisioning by a common invertebrate endosymbiont, *Wolbachia pipientis*, during periods of nutritional stress. PLoS Pathog 5:e1000368. doi:10.1371/journal.ppat.1000368

Brutinel ED, Dean AM, Gralnick JA (2013) Description of a riboflavin biosynthetic gene variant prevalent in the phylum proteobacteria. J Bacteriol 195:5479–5486

Chávez AS, Felsheim RF, Kurtti TJ, Ku PS, Brayton KA, Munderloh UG (2012) Expression patterns of *Anaplasma marginale* Msp2 variants change in response to growth in cattle, and tick cells versus mammalian cells. PLoS One 7(4):e36012. doi:10.1371/journal.pone.0036012

Comandatore F, Sassera D, Montagna M, Kumar S, Darby A, Blaxter M et al (2013) Phylogenomics and analysis of shared genes suggest a single transition to mutualism in *Wolbachia* of nematodes. Genome Biol Evol 5:1668–1674

Cordaux R, Michel-Salzat A, Bouchon D (2001) *Wolbachia* infection in crustaceans: novel hosts and potential routes for horizontal transmission. J Evol Biol 14:237–243

Cordaux R, Pichon S, Ling A, Perez P, Delaunay C, Vavre F, Bouchon D, Greve P (2008) Intense transpositional activity of insertion sequences in an ancient endosymbiont. Mol Biol Evol 25:1889–1895. doi:10.1093/molbev/msn134

Crossley RA, Gaskin DJ, Holmes K, Mulholland F, Wells JM, Kelly DJ, van Vliet AH, Walton NJ (2007) Riboflavin biosynthesis is associated with assimilatory ferric reduction and iron acquisition by *Campylobacter jejuni*. Appl Environ Microbiol 73:7819–7825

Dedeine F, Bandi C, Bouletreau M, Kramer L (2003) Insights into *Wolbachia* obligatory symbiosis. In: Bourtzis K, Miller TA (eds) Insect symbiosis. CRC Press, Florida, pp 267–282

Dell A, Galadari A, Sastre F, Hitchen P (2010) Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. Int J Microbiol. doi:10.1155/2010/148178

Fallon AM (2008) Cytological properties of an *Aedes albopictus* mosquito cell line infected with *Wolbachia* strain wAlbB. In Vitro Cell Dev Biol Anim 44:154–161. doi:10.1007/s11626-008-9090-4

Fallon AM, Baldridge GD, Higgins LA, Witthuhn BA (2013) *Wolbachia* from the planthopper *Laodelphax striatellus* establishes a robust, persistent, streptomycin-resistant infection in clonal mosquito cells. In Vitro Cell Dev Biol Anim 49:66–73. doi:10.1007/s11626-012-9571-3

Fallon AM, Baldridge GD, Carroll EM, Kurtz CM (2014) Depletion of host cell riboflavin reduces *Wolbachia* in cultured mosquito cells. In Vitro Cell Dev Biol Anim 50:707–713. doi:10.1007/s11626-014-9758-x

Felix C, Pichon S, Braquart-Varnier C, Braig H, Chen L, Garrett RA, Martin G, Greve P (2008) Characterization and transcriptional analysis of two gene clusters for type IV secretion machinery in *Wolbachia* of *Armadillidium vulgare*. Res Microbiol 159:481–485

Floate KD, Coghlin PC, Dosdall L (2011) A test using *Wolbachia* bacteria to identify Eurasian source populations of Cabbage Seed Pod Weevil, *Ceutorhynchus obstrictus* (Marsham), in North America. Environ Entomol 40:818–2011. doi:10.1603/EN10315

Gill AG, Darby AC, Makepeace BL (2014) Iron necessity: the secret of *Wolbachia's* success? PLoS Negl Trop Dis 16:e3224. doi:10.1371/journal.pntd.0003224

Gillespie JJ, Ammerman NC, Dreher-Lesnick SM, Rahman MS, Worley MJ, Setubal JC, Sobral BS, Azad AF (2009) An anomalous type IV secretion system in *Rickettsia* is evolutionarily conserved. PLoS One 4:e4833. doi:10.1371/journal.pone.0004833

Gillespie JJ, Brayton KA, Williams KP, Quevado Diaz MA, Brown WC, Azad AF, Sobral BW (2010) Phylogenomics reveals a diverse Rickettsiales type IV secretion system. Infect Immun 78:1809–1823

Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH (2008) How many species are infected with *Wolbachia*? A statistical analysis of current data. FEMS Microbiol Lett 281:215–220

Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T (2010) *Wolbachia* as a bacteriocyte-associated nutritional mutualist. Proc Natl Acad Sci USA 107:769–774

Hotopp JC, Lin M, Madupu R, Crabtree SV, Angiloui SV et al (2006) Comparative genomics of emerging human ehrlichiosis agents. PLoS Genet 2:e21

James AC, Dean MD, McMahon ME, Ballard JW (2002) Dynamics of double and single *Wolbachia* infections in *Drosophila simulans* from New Caledonia. Heredity 88:182–189

Jin C, Barrientos A, Tzagoloff A (2003) Yeast dihydroxybutanone phosphate synthase, an enzyme of the riboflavin biosynthetic pathway, has a second unrelated function in expression of mitochondrial respiration. J Biol Chem 278:14698–14703

Kambris Z, Cook PE, Phuc HK, Sinkins SP (2009) Immune activation by life-shortening *Wolbachia* and reduced filarial competence in mosquitoes. Science 326:134–136. doi:10.1126/science.1177531

Klasson L, Westberg J, Sapountzis P, Naslund K, Lutnaes Y, Darby AC, Veneti Z, Chen L, Braig HR, Garret R et al (2009) The mosaic structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*. Proc Natl Acad Sci USA 106:5725–5730

Knegt FH, Mello LV, Reis FC, Santos MT, Vicentini R, Ferraz LF, Ottoboni LM (2008) ribB and ribBA genes from *Acidithiobacillus ferrooxidans*: expression levels under different growth conditions and phylogenetic analysis. Res Microbiol 159:423–431

Koebnik R, Locher KP, Van Gelder P (2000) Structure and function of bacterial outer membrane proteins: barrels in a nutshell. Mol Microbiol 37:239–253

Kremer N, Voronin D, Charif D, Mavingui P, Mollereau B, Vavre F (2009) *Wolbachia* interferes with ferritin expression and iron metabolism in insects. PLoS Pathog 5:e1000630. doi:10.1371/journal.ppat.1000630

Li Z, Carlow CKS (2012) Characterization of transcription factors that regulate the type IV secretion system and riboflavin biosynthesis in *Wolbachia* of *Brugia malayi*. PLoS One 7:e51597. doi:10.1371/journal.pone0051597

Liu H, Bao W, Lin M, Niu H, Rikihisa Y (2012) *Ehrlichia* type IV secretion effector ECH0825 is translocated to mitochondria and curbs ROS and apoptosis by upregulating MnSOD. Cell Microbiol 14:1037–1050. doi:10.1111/j.1462-5822.2012.01775.x

Lo N, Paraskevopoulos C, Bourtzis K, O'Neill SL, Werren JH, Bordenstein SR, Bandi C (2007) Taxonomic status of the intracellular bacterium *Wolbachia pipientis*. Int J Syst Evol Microbiol 57:654–657

Lockwood S, Voth DE, Brayton KA, Beare PA, Brown WC, Heinzen RA, Broschat SL (2011) Identification of *Anaplasma marginale* type IV secretion system effector proteins. PLoS One 6:e27724. doi:10.1371/journal.pone.0027724

Masui S, Sasaki T, Ishikawa H (2000) Genes for the type IV secretion system in an intracellular symbiont, *Wolbachia*, a causative agent of various sexual alterations in arthropods. J Bacteriol 182:6529–6531

McDonagh B, Reguejo R, Fuentes-Almagro CA, Ogueta S, Bárcena JA, Padilla CA (2011) Thiol redox proteomics identifies differential targets of cytosolic and mitochondrial glutaredoxin-2 isoforms in *Saccharomyces cerevisiae*. Reversible S-glutathionylation of DHBP synthase (RIB3). J Proteomics 74:2487–2497

Melnikow E, Xu S, Liu J, Li L, Oksov Y, Ghedin E et al (2011) Interaction of a Wolbachia WSP-like protein with a nuclear-encoded protein of *Brugia malayi*. Int J Parasitol 41:1053–1061

Morrow JL, Frommer M, Shearman DC, Riegler M (2014) Tropical tephritid fruit fly community with high incidence of shared *Wolbachia* strains as platform for horizontal transmission of endosymbionts. Environ Microbiol 16:3622–3637. doi:10.1111/1462-2920.12382

Newton LG, Bordenstein SR (2011) Correlations between bacterial ecology and mobile DNA. Curr Microbiol 62:198–208. doi:10.1007/s00284-010-9693-3

Niu H, Kozjak-Pavlovic V, Rudel T, Rikihisa Y (2010) *Anaplasma phagocytophilum* Ats-1 is imported into host cell mitochondria and interferes with apoptosis induction. PLoS Pathog 6:e1000774

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

Noda H, Miyoshi T, Zhang Q, Watanabe K, Deng K, Hoshizaki S (2001b) *Wolbachia* infection shared among planthoppers (Homoptera: Delphacidae) and their endoparasite (Strepsiptera: Elenchidae): a probable case of interspecies transmission. Mol Ecol 10:2101–2106

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

Nothaft H, Szymanski CM (2010) Protein glycosylation in bacteria: sweeter than ever. Nat Rev 8:775–778. doi:10.1038/nrmicro2383

Ofran Y, Rost B (2006) ISIS: interaction sites identified from sequence. Bioinformatics 23:e13–e16. doi:10.1093/bioinformatics/bt303

Ohashi N, Zhi N, Lin Q, Rikihisa Y (2002) Characterization and transcriptional analysis of gene clusters for a type IV secretion machinery in human granulocytic and monocytic ehrlichiosis agents. Infect Immun 70:2128–2138

O'Neill SL, Pettigrew MM, Sinkins SP, Braig HR, Andreadis TG, Tesh RB (1997) In vitro cultivation of *Wolbachia pipientis* in an *Aedes albopictus* cell line. Insect Mol Biol 6:33–39

Pan X, Zhou G, Wu J, Bian G, Lu P, Raikhel AS, Xi Z (2012) *Wolbachia* induces reactive oxygen species (ROS)-dependent activation of the toll pathway to control dengue virus in the mosquito *Aedes aegypti*. Proc Natl Acad Sci USA 109:E23–E31. doi:10.1073/pnas.1116932108

Perrot-Minnot MJ, Guo LR, Werren JH (1996) Single and double infections with *Wolbachia* in the parasitic wasp *Nasonia vitripennis*: effects on compatibility. Genetics 143:961–972

Pichon S, Bouchon D, Cordaux R, Chen L, Garret RA, Greve P (2009) Conservation of the type IV secretion system throughout *Wolbachia* evolution. Biochem Biophys Res Commun 385:557–562

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

Raychoudhury R, Baldo L, Oliveira DCSG, Werren JH (2008) Modes of acquisition of *Wolbachia*: horizontal transfer, hybrid introgression, and convergence in the *Nasonia* species complex. Evolution 63:165–183. doi:10.1111/j.1558-5646.2008.00533.x

Rio R, Hu Y, Aksoy S (2004) Strategies of the home-team: symbioses exploited for vector-borne disease control. Trends Microbiol 12:325–336. doi:10.1016/j.tim.2004.05.001

Sanogo YO, Dobson SL, Bordenstein SR, Novak RJ (2007) Disruption of the *Wolbachia* surface protein gene wspB by a transposable element in mosquitoes of the *Culex pipiens* complex (Diptera, Culicidae). Insect Mol Biol 16:143–154

Saridaki A, Bourtzis K (2010) *Wolbachia*: more than just a bug in insect genitals. Curr Opin Microbiol 13:67–72

Sarkar M, Troese MJ, Kearns SA, Yang T, Reneer DV, Carlyon JA (2008) *Anaplasma phagocytophilum* MSP2(P44)-18 predominates and is modified into multiple isoforms in human myeloid cells. Infect Immun 76:2090–2098. doi:10.1128/IAI.01594-07

Schuler H, Bertheau C, Egan SP, Feder JL, Riegler M, Schlick-Steiner BC, Steiner FM, Johannesen J, Kern P, Tuba K, Lakatos F, Koppler K, Arthofer W, Stauffer C (2013) Evidence for a recent horizontal transmission and spatial spread of *Wolbachia* from endemic *Rhagoletis cerasi* (Diptera: Tephritidae) to invasive *Rhagoletis cingulata* in Europe. Mol Ecol 22:4101–4111. doi:10.1111/mec.12362

Shih KM, Gerenday A, Fallon AM (1998) Culture of mosquito cells in Eagle's medium. In Vitro Cell Dev Biol Anim 34:629–630

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Sodung J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. doi:10.1038/msb.2011.75

Sinkins SP, Gould F (2006) Gene drive systems for insect disease vectors. Nat Rev Genet 7:427–435

Strubing U, Lucius R, Hoerauf A, Pfarr KM (2010) Mitochondrial genes for heme-dependent respiratory chain complexes are up-regulated after depletion of *Wolbachia* from filarial nematodes. Int J Parasitol 15:1193–1202. doi:10.1016/j.ijpara.2010.03.004

Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland

Taylor MJ, Bandi C, Hoerauf A (2005) *Wolbachia* bacterial endosymbionts of filarial nematodes. Adv Parasitol 60:245–284

Uday J, Puttaraju HP (2012) Comparative analysis of *Wolbachia* surface protein in *D. melanoagster*, *A. tabida* and *B. malayi*. Bioinformation 8(15):711–715. doi:10.6026/97320630008711

Werren JH, Baldo L, Clark ME (2008) *Wolbachia*: master manipulators of invertebrate biology. Nat Rev Microbiol 6:741–751

Worst DJ, Gerrits MM, Vandenbroucke-Grauls CM, Kusters JG (1998) *Helicobacter pylori* ribBA-mediated riboflavin production is involved in iron acquisition. J Bacteriol 180:1473–1479

Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw E, Martin W, Esser C et al (2004) Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol 2:0327–0341

Wu B, Novelli J, Foster J, Vaisvila R, Conway L, Ingram J, Ganatra M, Rao AU, Hamza I, Slatko B (2009) The heme biosynthetic pathway of the obligate *Wolbachia* endosymbiont of *Brugia malayi* as a potential anti-filarial drug target. PLoS Negl Trop Dis 3:e475. doi:10.1371/journal.pntd.0000475

Yang C-H, Xiao J-H, Niu L-M, Ma G-C, Cook JM, Bian S-N, Fu Y-G, Huang D-H (2012) Chaos of *Wolbachia* sequences inside the compact fig syconia of *Ficus benjamina* (Ficus: Moraceae). PLoS One. doi:10.1371/journal.pone.0048882

Yurgel SN, Rice J, Domreis E, Lynch J, Sa N, Qamar Z, Rajamani S, Gao M, Roje S, Bauer WD (2014) *Sinorhizobium meliloti* flavin secretion and bacteria-host interaction: role of the bifunctional RibBA protein. Mol Plant-Microbe Interact 27:437–445. doi:10.1094/MPMI-11-13-0338-R

Zabalou S, Riegler M, Theodorakopoulou M, Stauffer C, Savakis C, Bourtzis K (2004) *Wolbachia*-induced cytoplasmic incompatibility as a means for insect pest population control. Proc Natl Acad Sci USA 101:15042–15045. doi:10.1073/pnas.0403853101

Zechner EL, Lang S, Schildbach JF (2012) Assembly and mechanisms of bacterial type IV secretion machines. Phil Trans R Soc B 367:1073–1087. doi:10.1098/rstb.2011.0207

Zhang KJ, Han X, Hong XY (2013) Various infection status and molecular evidence for horizontal transmission and recombination of *Wolbachia* and *Cardinium* among rice planthoppers and related species. Insect Sci 3:329–344. doi:10.1111/j.1744-7917.2012.01537.x

Zug R, Hammerstein P (2014) Bad guys turned nice? A critical assessment of *Wolbachia* mutualisms in arthropod hosts. Biol Rev. doi:10.1111/brv.12098

Zug R, Koehncke A, Hammerstein P (2012) Epidemiology in evolutionary time: the case of *Wolbachia* horizontal transmission between arthropod species. J Evol Biol 25:2149–2160. doi:10.1111/j.1420-9101.2012.02601.x