Evolutionary Genetics of Cytoplasmic Incompatibility Genes *cifA* and *cifB* in Prophage WO of *Wolbachia*

Amelia R.I. Lindsey[1], Danny W. Rice[2], Sarah R. Bordenstein[3], Andrew W. Brooks[3,4], Seth R. Bordenstein[3,4,5,6,]*, and Irene L.G. Newton[2,]*

[1]Department of Entomology, University of California Riverside

[2]Department of Biology, Indiana University, Bloomington

[3]Department of Biological Sciences, Vanderbilt University

[4]Vanderbilt Genetics Institute, Vanderbilt University

[5]Vanderbilt Institute for Infection, Immunology and Inflammation, Vanderbilt University

[6]Department of Pathology, Microbiology and Immunology, Vanderbilt University

*Corresponding authors: E-mails: s.bordenstein@vanderbilt.edu; irnewton@indiana.edu.

Abstract

The bacterial endosymbiont *Wolbachia* manipulates arthropod reproduction to facilitate its maternal spread through host populations. The most common manipulation is cytoplasmic incompatibility (CI): *Wolbachia*-infected males produce modified sperm that cause embryonic mortality, unless rescued by embryos harboring the same *Wolbachia*. The genes underlying CI, *cifA* and *cifB*, were recently identified in the eukaryotic association module of *Wolbachia*'s prophage WO. Here, we use transcriptomic and genomic approaches to address three important evolutionary facets of the *cif* genes. First, we assess whether or not *cifA* and *cifB* comprise a classic toxin–antitoxin operon in *w*Mel and show that the two genes exhibit striking, transcriptional differences across host development. They can produce a bicistronic message despite a predicted hairpin termination element in their intergenic region. Second, *cifA* and *cifB* strongly coevolve across the diversity of phage WO. Third, we provide new domain and functional predictions across homologs within *Wolbachia*, and show that amino acid sequences vary substantially across the genus. Finally, we investigate conservation of *cifA* and *cifB* and find frequent degradation and loss of the genes in strains that no longer induce CI. Taken together, we demonstrate that *cifA* and *cifB* exhibit complex transcriptional regulation in *w*Mel, provide functional annotations that broaden the potential mechanisms of CI induction, and report recurrent erosion of *cifA* and *cifB* in non-CI strains, thus expanding our understanding of the most widespread form of reproductive parasitism.

Key words: symbiosis, reproductive manipulation, gene loss, bacteriophage, prophage.

Introduction

The genus *Wolbachia* is the most widespread group of maternally transmitted endosymbiotic bacteria (Zug et al. 2012). They occur worldwide in numerous arthropods and nematodes and can selfishly manipulate reproduction (Werren et al. 2008), confer antiviral defense (Teixeira et al. 2008; Bian et al. 2010), and assist reproduction and development of their hosts (Hoerauf et al. 1999; Dedeine et al. 2001; Hosokawa et al. 2010). The most common parasitic manipulation is cytoplasmic incompatibility (CI), whereby *Wolbachia*-infected males

produce modified sperm that can only be rescued by eggs infected with the same *Wolbachia* strain (Yen and Barr 1971). If the modified sperm fertilize eggs infected with no *Wolbachia* (unidirectional CI) or a genetically incompatible *Wolbachia* strain (bidirectional CI), then delayed histone deposition, improper chromosome condensation, and cell division abnormalities result in embryonic arrest and death (Lassy and Karr 1996; Tram and Sullivan 2002; Serbus et al. 2008; Landmann et al. 2009). Other described reproductive manipulations include parthenogenesis (Stouthamer et al. 1990),

male-killing (Hurst et al. 1999), and feminization (Rousset et al. 1992), all of which enhance the fitness of *Wolbachia*-infected females and assist the spread of the infected matri-line through a population. These manipulations, once sustained, can also impact host evolution including speciation (Bordenstein et al. 2001; Jaenike et al. 2006; Brucker and Bordenstein 2013) and mating behaviors (Randerson et al. 2000; Moreau et al. 2001; Miller et al. 2010; Shropshire and Bordenstein 2016).

In addition to the aforementioned reproductive manipulations, *Wolbachia* strains affect host biology by provisioning nutrients (Hosokawa et al. 2010), altering host survivorship (Min and Benzer 1997) and fecundity (Stouthamer and Luck 1993; Dedeine et al. 2001), and importantly, protecting the host against pathogens (Teixeira et al. 2008; Kambris et al. 2009; Moreira et al. 2009; Bian et al. 2010; Hughes et al. 2011; Walker et al. 2011). The combination of reproductive manipulations that enable *Wolbachia* to spread in a population and the ability to reduce vector competence through pathogen protection have placed *Wolbachia* in the forefront of efforts to control disease carrying arthropod populations (Turelli and Hoffmann 1991; Zabalou et al. 2004; Hoffmann et al. 2011; Walker et al. 2011; LePage and Bordenstein 2013; Bourtzis et al. 2014). Despite these important applications, the widespread prevalence of *Wolbachia* across arthropod taxa (Werren and Windsor 2000; Hilgenboecker et al. 2008; Zug et al. 2012), and decades of research, only recently have the genes underlying CI been determined (Beckmann et al. 2017; LePage et al. 2017).

Two studies converged on the same central finding: Coexpression of a pair of syntenic genes recapitulates the CI phenotype (Beckmann et al. 2017; LePage et al. 2017). Uninfected *Drosophila melanogaster* males transgenically expressing the two genes from *w*Mel *Wolbachia* caused CI-like embryonic lethality when crossed with uninfected females that was notably rescued by *w*Mel-infected females (LePage et al. 2017). Additionally, the two *w*Mel genes separately enhanced *w*Mel-induced CI in a dose-dependent manner when expressed in infected males, and the CI was again rescued by *w*Mel-infected females (LePage et al. 2017). In the other study, CI-like embryonic lethality was also recapitulated in *D. melanogaster* males through transgenic coexpression of homologous transgenes *cidA* and *cidB*, encoded by the *w*Pip strain of *Wolbachia* that naturally infect *Culex* mosquitoes (Beckmann et al. 2017). These two genes are located in the recently discovered eukaryotic association module of temperate phage WO (Bordenstein SR and Bordenstein SR 2016), which was previously implicated in influencing CI (Masui et al. 2000; Sinkins et al. 2005; Bordenstein et al. 2006; Duron et al. 2006). The presence of these genes within prophage WO has implications for the transmission of these genes because temperate phage WO exhibits frequent lateral transfers between *Wolbachia* (Bordenstein and Wernegreen 2004; Chafee et al. 2010) while *Wolbachia* are mainly

vertically transmitted from mothers to offspring. The genes were proposed as candidate CI effectors due to the presence of one of the protein products in the spermathecae of infected female mosquitoes (Beckmann and Fallon 2013) and their absence in the *w*Au *Wolbachia* strain that lost CI function (Sutton et al. 2014).

The *w*Mel homologs of these genes are designated CI factors *cifA* (locus WD0631) and *cifB* (locus WD0632), with *cifA* always encoded directly upstream of *cifB* (LePage et al. 2017). The gene set occurs in varying copy number across 11 total CI-inducing strains, and the copy number tentatively correlates with CI levels. Core sequence changes of the two genes exhibit a pattern of codivergence and in turn closely match bidirectional incompatibility patterns between *Wolbachia* strains. Homologs of CifA and CifB protein sequences belong to four distinct phylogenetic Types (designated Types I–IV) that do not correlate with various phylogenies of *Wolbachia* housekeeping genes or *gpW* (locus WD0640) in phage WO (LePage et al. 2017). The homologous sequences in *w*Pip also cluster in Type I, though they are 66% and 76% different from *w*Mel's, respectively (Beckmann et al. 2017). Hereinafter we use *cifA* and *cifB* to refer to these genes, unless specifically referring to analyses of the *w*Pip homologs, *cidA* and *cidB*. In vitro functional analyses revealed that *cidB* encodes deubiquitylase activity, and *cidA* encodes a protein that binds CidB (Beckmann et al. 2017). Mutating the predicted catalytic residue in the deubiquitylating domain of CidB results in a loss of the CI-like function in transgenic flies (Beckmann et al. 2017). Whether these genes or other alleles have additional enzymatic or regulatory roles and which other residues are important for function remain open questions.

There are important considerations for the location, organization, and characterization of these genes. Whether or not *cifA* and *cifB* form a strict, toxin–antitoxin operon is debatable, and likewise has important implications for how gene expression is regulated by *Wolbachia* during host infection. Support for the operon hypothesis is based on weak transcription across the junction between *cidA* and *cidB*, inferred to be due to the presence of bicistronic mRNA (Beckmann and Fallon 2013; Beckmann et al. 2017); an alternative explanation is transcriptional slippage. Quantitative transcription analyses and various computational predictions of operon structure do not support the operon hypothesis (LePage et al. 2017). Moreover and importantly, transgenic studies show that both *cifA* and *cifB* are required for induction of CI and thus cannot form a strict toxin (*cifB*)–antitoxin (*cifA*) system as both genes positively contribute to CI and can individually enhance *Wolbachia*-induced CI (LePage et al. 2017). However, like toxin–antitoxin systems, CidA binds CidB in vitro and expression of *cidA* rescues temperature-sensitive growth inhibition induced by *cidB* expression in *Saccharomyces*, via an as-yet-unknown mechanism (Beckmann et al. 2017).

As it stands now, the genes remain largely unannotated with the exception of a few small domains. If other predicted protein domains occur in CifA and CifB, they could allow for new hypotheses for the mechanism of CI. Finally, the sequence diversity and/or loss of *cif* genes across the *Wolbachia* tree may give insights into the selective conditions that maintain the *cif* genes versus those that do not. Exploration of *cif* gene regulation, expression, and function thus can provide a framework for more targeted investigations of *Wolbachia*–host interactions, and potentially inform the deployment of *Wolbachia*-based arthropod control.

Materials and Methods

Expression

For analysis of RNAseq data, we used our published approach (Gutzwiller et al. 2015). Briefly, fastq sequences for 1-day-old male and female flies were mapped against the *Wolbachia* *w*Mel reference genome (GenBank AE017196) using bwa mem v. 0.7.5a with default parameters in paired-end mode. Mapped reads were sorted and converted to BAM format using samtools v0.1.19 after which BAM files were used as input to Bedtools (bedcov) to generate pileups and count coverage at each position. For expression correlations between genes, the raw RNAseq counts were divided by (gene length $+ 99$), where 99 corresponds to read length $(100) - 1$. Within a growth stage these values were multiplied by $1e^6/(\text{sum of}$ values in stage) (Li and Dewey 2011). A pairwise distance between all genes was defined as $(1 - R)$, where the R is the Pearson correlation coefficient between the normalized expression values of two genes. Possible negative correlations would be "penalized" here, resulting in a larger distance. Distances were clustered using the Kitsch program of PHYLIP (Felsenstein 1989).

Operon Prediction In Silico

We used the dynamic profile of the transcriptome above to identify operons within the *w*Mel genome using two different approaches. We used the program Rockhopper (McClure et al. 2013), with default parameters, in conjunction with the BAM files generated above to delineate likely operons across the entire genome. The Arnold web server (http://rna.igmors.u-psud.fr/toolbox/arnold/) was used to predict hairpin transcription termination elements (Gautheret and Lambert 2001; Macke et al. 2001).

Nucleic Acid Extractions and Quantitative Reverse Transcription Polymerase Chain Reaction

To identify *Wolbachia* gene expression in adult male and female *D. melanogaster*, RNA was extracted from individual, age-matched flies (1–3 days old, stock 145) using a modified Trizol extraction protocol. Briefly, 500 μl of Trizol was added to individual flies and samples homogenized using a pestle. After a 5-min incubation at room temperature, a 12,000 rcf centrifugation (at 4 °C for 10 min) was followed by a chloroform extraction. Aqueous phase containing RNA was extracted a second time with phenol: Chloroform before isopropanol precipitation of RNA. This RNA pellet was washed and resuspended in THE RNA Storage Solution (Ambion). RNA used in subsequent analyses was subjected to a short DNAse treatment (10 min at 37 °C then 10 min at 75 °C to inactivate the enzyme). To detect the number of *cifA* and *cifB* transcripts as well as RNA levels across the junction between *cifA* and *cifB*, we utilized the RNA extracted from these flies and the SensiFAST SYBER Hi-ROX One-step RT mix (Bioline) and the Applied Biosystems StepOne Real-time polymerase chain reaction (PCR) system. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) was performed with the following primer sets: *cifAF*: ATAAAGGCGTTTCAGCAGGA, *cifAR*: AGCAAAGCGTTCACATTTCC *cifBF*: TACGGGAAG TTTCATGCACA, *cifBR*: TTGCCAGCCATCATTCATAA; *cifA*_endF: TCTGGTTCTCATAAGAAAAGAAGAATC, *cifB*_begR: AACCATCAAGATCTCCATCCA. As a reference for transcription activity of the core *Wolbachia* genome, we utilized the *Wolbachia* *ftsZ* gene (forward: TTTTGTTGTCGCAAATACCG; reverse: CCATTCCTGCTGTGATGAAA). We designed primers to *ftsZ* because as a core protein involved in cell division, the quantities of *ftsZ* would better correlate with bacterial numbers and activity. Reactions were performed in duplicate or triplicate in a 96-well plate and CT values generated by the machine were used to calculate the relative amounts of *Wolbachia* using the $\Delta\Delta$Ct (Livak) method.

To identify a bicistronic message encompassing *cifA* and *cifB*, we designed primers based on the 5′-region of *cifA* and the 3′-region of *cifB* (WD0631F: ATAAAGGCGTTTC AGCAGGA; WD0632R: TTGCCAGCCATCATTCATAA). We extracted RNA from whole animals and performed a DNAse treatment (as described above) before using the iScript first strand synthesis kit (Biorad) to generate cDNA. Negative controls included RT minus reactions. Resulting cDNA and negative controls were used in PCR reactions with the primers above and the following cycling conditions: 95 °C 5 min then 35 cycles of 95 °C for 1 min, 64 °C for 1 min, 72 °C for 2.5 min followed by a final extension of 72 °C for 10 min and using the HF Phusion enzyme mix (NEB). As a positive control, to confirm that we could amplify long mRNAs from these samples, we used the 16S rRNA gene primers 27F (AGAGTTTGATCCTGGCTCAG) and 1492R (GGTTACCTTGTTACGACTT) with the same cycling conditions as above except that the annealing temperature was 55 °C.

Correlated Cif Trees and Distance Matrices

Quantifying congruence scores between the CifA and CifB trees was carried out with Matching Cluster (MC) and

Robinson–Foulds (RF) metrics using a custom python script previously described (Brooks et al. 2016) and the TreeCmp program (Bogdanowicz et al. 2012). MC weights topological congruency of trees, similar to the widely used RF metric. However, MC takes into account sections of subtree congruence and therefore is a more refined evaluation of small topological changes that affect incongruence. Significance in the MC and RF analyses was determined by the probability of 100,000 randomized bifurcating dendrogram topologies yielding equivalent or more congruent trees than the actual tree. Normalized scores were calculated as the MC and RF congruency score of the two topologies divided by the maximum congruency score obtained from random topologies. The number of trees that had an equivalent or better score than the actual tree was used to calculate the significance of observing that topology. Mantel tests were also performed on the CifA and CifB patristic distance matrices calculated in Geneious v8.1.9 (Kearse et al. 2012). A custom Jupyter notebook (Pérez and Granger 2007) running python v3.5.2 (http://python.org) was written in the QIIME2 (Caporaso et al. 2010) anaconda environment. The Mantel test (Mantel 1967) utilized the scikit-bio v0.5.1 (scikit-bio.org) Mantel function run, using scikit-bio distance matrix objects for each gene. The Mantel test was run with 100,000 permutations to calculate significance of the Pearson correlation coefficient between the two matrices using a two-sided correlation hypothesis.

Genomes Used in Comparative Analyses

In order to identify *cif* homologs across the *Wolbachia* genomes, we defined orthologs across existing, sequenced genomes using reciprocal best BlastP. We included *Wolbachia* genomes across five supergroups: Monophyletic clades of *Wolbachia* based on housekeeping genes, denoted by uppercase letters (O'Neill et al. 1992; Werren et al. 1995). Supergroups A and B are the major arthropod infecting lineages, whereas C and D infect nematodes (Bandi et al. 1998). Supergroup F *Wolbachia* infect a variety of hosts (Lo et al. 2002). Included in this analysis were nine type A strains (*w*Ri, *w*Suzi, *w*Ha, *w*Mel, *w*MelPop, *w*Au, *w*Rec, *w*Uni, and *w*VitA), seven type B strains (*w*PipJHB, *w*PipPel, *w*Bol1-b, *w*No, *w*Tpre, *w*AlbB, and *w*Di), two type C strains (*w*Ov and *w*Oo), and one each type D (*w*Bm) and type F (*w*Cle). We included all genomic data available for each strain such that if multiple assemblies existed for each *Wolbachia* variant (such as in the case of *w*Uni) we included the union of all available contigs for that strain. *Wolbachia* orthologs were defined based on reciprocal best blast hits between amino acid sequences in *Wolbachia* genomes. An orthologous group of genes was defined by complete linkage such that all members of the group had to be the reciprocal best hit of all other members of the group. Information on strain phenotypes, hosts, and accession numbers can be found in table 1.

Cif Phylogenetics

CifA and CifB protein sequences were identified using BlastP searches of WOMelB WD0631 (NCBI accession number AAS14330.1) and WD0632 (AAS14331.1), respectively. Homologs were selected based on: 1) $E = \leq 10^{-30}$, 2) query coverage greater than 70%, and 3) presence in fully sequenced *Wolbachia* genomes. All sequences were intact with the exception of a partial WOSuziC CifA (WP_044471252.1) protein. The missing N-terminus was translated from the end of contig accession number CAOU02000024.1 and concatenated with partial protein WP_044471252.1 for analyses, resulting in 100% amino acid identity to WORiC CifA (WP_012673228.1). In addition, two previously identified sequences (LePage et al. 2017), WORecB CifB and WORiB CifB, were not available in NCBI's database and translated from nucleotide accession numbers JQAM01000018.1 and CP001391.1, respectively. The previously identified WOSol homologs (CifA: AGK87106 and CifB: AGK87078) (LePage et al. 2017) were also included in our analyses. All protein sequences were aligned with the MUSCLE (Edgar 2004) plugin in Geneious Pro version 8.1.7 (Kearse et al. 2012); the best models of evolution, according to corrected Akaike (Hurvich and Tsai 1993) information criteria, were estimated to be JTT-G using the ProtTest server (Abascal et al. 2005); and phylogenetic trees were built using the MrBayes (Ronquist et al. 2012) plugin in Geneious.

Protein Structure

All candidate CI gene protein sequences were individually assessed for the presence of domain structure using HHpred (https://toolkit.tuebingen.mpg.de/hhpred/; Söding et al. 2005)) with default parameters and the following databases: SCOPe70 (v.2.06), Pfam (v.31.0), SMART (v6.0), and COG/KOG (v1.0). Schematics were created in inkscape (https://inkscape.org/), to show regions with significant structural hits, as determined by probabilities greater than 50%, or greater than 20% and in the top five hits.

Protein Conservation

Protein conservation was determined with the Protein Residue Conservation Prediction tool (http://compbio.cs.princeton.edu/conservation/index.html; Capra and Singh 2007), using aligned amino acid sequences, Shannon entropy scores, a window size of zero, and sequence weighting set to "false." Conservation was subsequently plotted in R version 3.3.2, and module regions were delineated according to the coordinates of the WOMelB modules within the alignment. CI gene conservation scores were calculated separately for Type I sequences, and for all Types together. For CifB Type I sequences, the WOVitA4 ortholog was left out, due to the extended C-terminus of that protein. Conservation scores were also calculated for "control proteins": Wsp (*Wolbachia* surface protein),

Table 1

Genomes Used in Comparative Analyses of *cifA* and *cifB*

Supergroup	Strain	Host	Reproductive Phenotypes	Accession Number
A	*w*Mel	*Drosophila melanogaster*	CI	NC_002978.6
	*w*MelPop	*Drosophila melanogaster*	CI	AQQE00000000.1
	*w*Rec	*Drosophila recens*	CI	NZ_JQAM00000000.1
	*w*Au	*Drosophila simulans*	None	LK055284.1
	*w*Ha	*Drosophila simulans*	CI	NC_021089.1
	*w*Ri	*Drosophila simulans*	CI	NC_012416.1
	*w*Suzi	*Drosophila suzukii*	None	NZ_CAOU00000000.2
	*w*Uni	*Muscidifurax uniraptor*	PI	NZ_ACFP00000000.1
	*w*VitA	*Nasonia vitripennis*	CI	NZ_MUJM00000000.1
B	*w*AlbB	*Aedes albopictus*	CI	CAGB00000000.1
	*w*No	*Drosophila simulans*	CI	NC_021084.1
	*w*Di	*Diaphorina citri*	Undetermined	NZ_KB223540.1
	*w*Tpre	*Trichogramma pretiosum*	PI	CM003641.1
	*w*VitB	*Nasonia vitripennis*	CI	AERW00000000.1
	*w*Bol1-b	*Hypolimnas bolina*	CI, MK	NZ_CAOH00000000.1
	*w*PipJHB	*Culex quinquefasciatus*	CI	ABZA00000000.1
	*w*PipPel	*Culex pipiens*	CI	NC_010981.1
C	*w*Oo	*Onchocerca ochengi*	OM	NC_018267.1
	*w*Ov	*Onchocerca volvulus*	OM	NZ_HG810405.1
D	*w*Bm	*Brugia malayi*	OM	NC_006833.1
F	*w*Cle	*Cimex lectularius*	OM	NZ_AP013028.1

Note.—Reproductive phenotypes include: CI, parthenogenesis-inducing (PI), male-killing (MK), obligate mutualism (OM), no phenotype discovered after assessment (None), and phenotype was not assayed (Undetermined).

known to be affected by frequent recombination events (Baldo et al. 2005), and FtsZ, which is relatively unaffected by recombination (Baldo, Dunning Hotopp, et al. 2006; Ros et al. 2009). Variation in amino acid conservation between modules and nonmodule regions was assessed in R version 3.3.2 with a one-way ANOVA including "region" (either the unique module number, or "nonmodule") as a fixed effect, and followed by Tukey Honest Significant Difference for post hoc testing.

Cif Modules

The WOMelB structural regions delineated by HHpred were used to search for the presence of Cifs or remnants of Cifs across the *Wolbachia* phylogeny. Amino acid sequences of the WOMelB modules were queried against complete genome sequences (table 1) using TBlastN. Any hit that was at least 40% of the length and 40% identity, or at least 90% of the length and 30% identity of the WOMelB module was considered a positive match. Module presence was plotted across a *Wolbachia* phylogeny constructed using the five multilocus sequence typing (MLST) genes defined by Baldo, Dunning Hotopp, et al. (2006). Nucleotide sequences were aligned with MAFFT version 7.271 (Katoh and Standley 2013), and concatenated prior to phylogenetic reconstruction with RAxML version 8.2.8 (Stamatakis 2014), the GTRGAMMA substitution model, and 1,000 bootstrap replicates. We also searched for *cif*-like regions in *Cardinium*: An

unrelated endosymbiont that can also cause CI in arthropods (Penz et al. 2012). Here, searches were performed with TBlastN and restricted to all available *Cardinium* sequence in NCBI GenBank (taxid: 273135).

Hidden Markov Model Searches

To identify *cif* homologs in draft *Wolbachia* genome assemblies we used the program suite HMMER (Eddy 2011). We defined *cif* Types based on our phylogenetic trees (fig. 4) and used aligned amino acids from these Types as input to HMMBUILD, using default parameters. We then searched six *Wolbachia* WGS assemblies (NCBI project numbers PRJNA310358, PRJNA279175, PRJNA322628) using HMMSEARCH with –F3 1e-20 –cut_nc and –domE 1e-10. Regardless of thresholds used, or *cif* type of HMM, resulting hits did not differ.

Results

cifA and *cifB* Are Cotranscribed but Differentially Regulated in *w*Mel

In bacteria, genes are commonly grouped into a single transcriptional unit under the control of one promoter, referred to as an operon. Because *cifA* and *cifB* are syntenic across prophage WO of *Wolbachia* and both involved in CI, we aimed to assess whether *cifA* and *cifB* are cotranscribed. We performed RT-PCR using primers that amplify the entire region from the

(A)

(B)
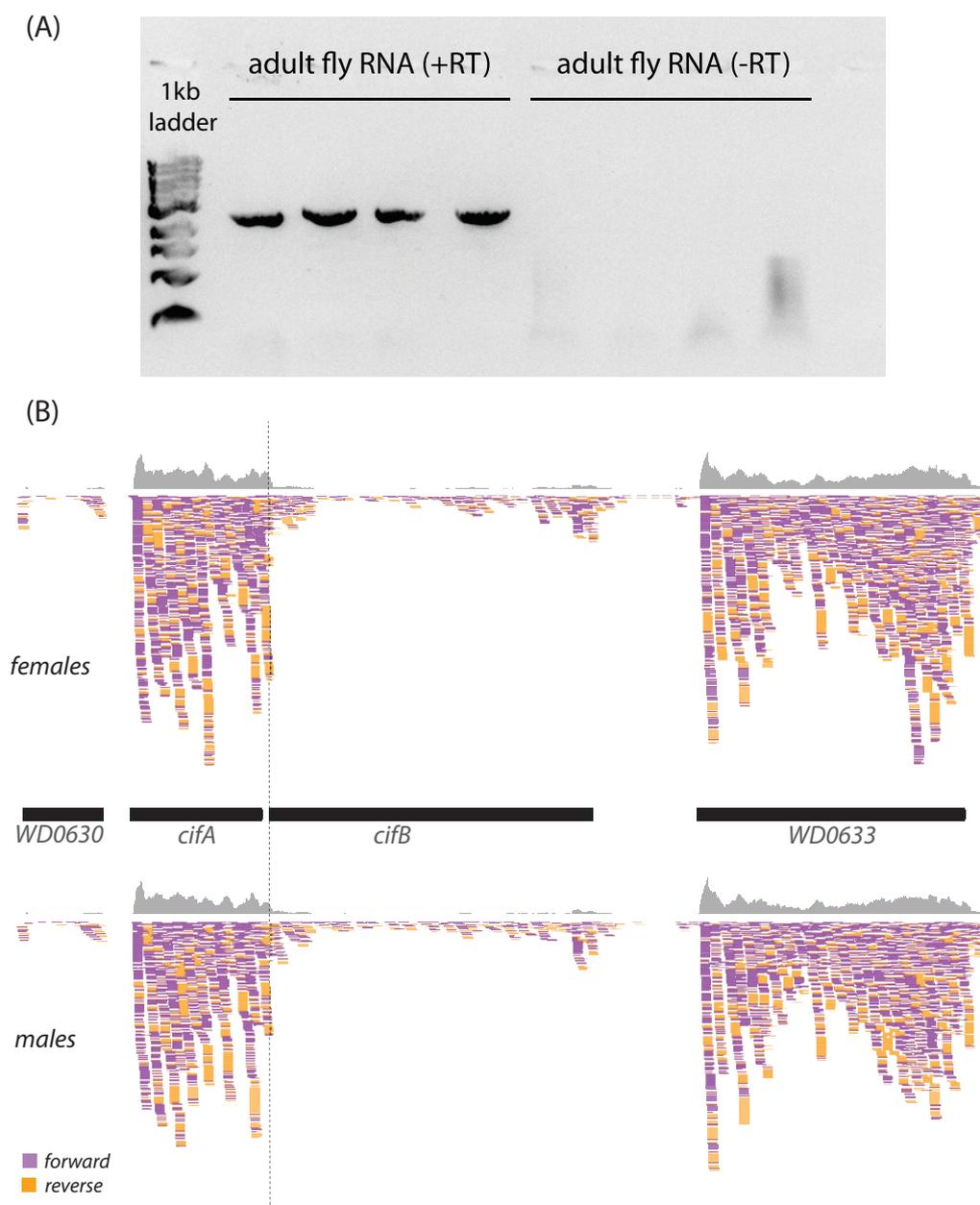
Fig. 1—Expression of *cifA* and *cifB* in adult flies. (*A*) Amplification of *cifA*–*cifB* bicistronic message from cDNA generated from adult flies. Positive amplification occurred in both male and female adult flies. RT minus controls included. (*B*) RNAseq expression from 1-day-old female and male *Drosophila melanogaster* flies. Raw reads were mapped to the *w*Mel assembly (using bwa) and coverage visualized using the Integrated Genomics Viewer (v2.3.77). The start of the *cifB* open reading frame is denoted by a vertical, dotted line.

start of *cifA* to the end of *cifB* (∼2.5 kb in total). cDNA amplification of the region from *w*Mel-infected male and female flies was successful (fig. 1*A*), and the transcript was confirmed to be *cifA*–*cifB* using Sanger sequencing from the forward and reverse ends of the cDNA amplicon (supplementary file S2, Supplementary Material online). We could not amplify a larger transcript from the loci flanking *cifA* and *cifB*, suggesting that the *cifA*–*cifB* transcript is a discrete unit.

Operons are often comprised of loci encoding related processes that can therefore be coregulated conveniently through the control of transcription from one promoter. To assess whether *cifA* and *cifB* are coregulated, we reasoned that strictly coregulated loci will have correlated gene expression across host development and similar total expression levels in whole animals. We therefore utilized an existing RNAseq data set for *Wolbachia* in *Drosophila melanogaster*, covering 24 life cycle stages and 3 time samplings each for adult males

and females (Gutzwiller et al. 2015). We mapped reads to the existing wMel assembly (see Materials and Methods) and calculated Pearson correlation coefficients of normalized expression values across host development for between all gene pairs. cifA is expressed at much higher absolute levels than cifB (fig. 1B; 8-fold higher based on RPKM values across both genes), and cifA and cifB expression is weakly, negatively correlated (Pearson r: −0.40; P-value: 0.014), suggesting that the expression level of one could have a negative influence on the level of the other. To confirm differential expression of cif genes in wMel, we performed a quantitative RT-PCR analysis of gene expression from 3-day-old male and female flies (fig. 2). We observed transcripts covering the junction between cifA and cifB. However, transcripts covering this junction were more similar to the expression levels in cifA, whereas expression of cifB was 9-fold less, supporting results from the RNAseq analysis.

As a possible explanation for the large absolute differences in transcription, we examined the intergenic sequence between cifA and cifB and identified a Rho-independent transcription terminator at nucleotides 618649–618668. This terminator region is predicted to form a GC-rich hairpin (50% GC compared with the Wolbachia wMel genome-wide 35%) in newly synthesized mRNA message proximal to the RNA polymerase. There are two explanations for how the terminator might explain the transcript abundance differences between cifA and cifB, and both have an impact on the operon hypothesis. First, cifA and cifB have their own promoters, but occasionally the genes are cotranscribed as a bicistronic message due to an imperfect hairpin terminator at the end of cifA. In this model, cifA and cifB do not form an operon. Alternatively, the cifA and cifB operon has a single promoter upstream of cifA, and the imperfect terminator provides a mechanism to control transcriptional differences between cifA and cifB in the operon. Functionally resolving whether cifA and cifB have the same or different promoters will be the ultimate arbiter of the two models.

In order to identify loci with similar expression patterns during host development, we clustered all wMel genes based on their similarity in expression across Drosophila development (supplementary fig. S1, Supplementary Material online). cifA did not group with cifB in wMel (fig. 3), suggesting that these two genes are not similarly expressed. Indeed, the pattern of cifA expression differs strikingly from that of cifB. For example, cifA is relatively highly expressed during late embryogenesis and in adults, whereas cifB is relatively highly expressed during the first two-thirds of embryogenesis, and during larval stages (fig. 3A). Curiously, the expression profile of cifA in flies during development is most closely correlated with the wsp locus WD1063 (fig. 3B).

Because of the dramatic absolute difference in cifA and cifB transcript levels, computational methods for operon prediction do not support their cotranscription. For example, after mapping reads to the wMel assembly, we used the resulting
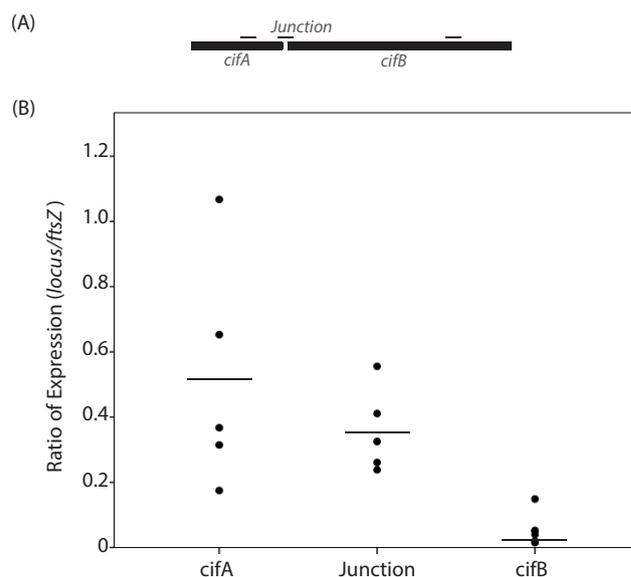
Fig. 2—Relative expression ratio of cifA, the junction between cifA/cifB, and cifB to ftsZ. Expression of both genes and their junction was quantified using qRT-PCR, and normalized to Wolbachia ftsZ gene expression. cifB gene expression is significantly less than that of the junction ($t = 3.220$, df = 16, $P = 0.005$) and less than cifA ($t = −3.840$, df = 17, $P = 0.001$).

BAM files as input to Rockhopper (McClure et al. 2013). The program was able to correctly identify known operons in wMel (such as the T4SS WD0004–WD0008 and the ribosomal protein operon), but it did not identify cifA and cifB as an operon.

In summary, although a bicistronic message of cifA and cifB was detected by qPCR, their absolute and relative expression levels are drastically different. A termination signal in their intergenic sequence may limit expression of the bicistronic message and could explain the much higher absolute level of cifA. Given the negative correlation across growth stages, some entity that activates cifA transcription, or a cifA product itself, could repress cifB transcription. Clearly, cifA and cifB are not a traditional operon and functionally resolving their promoter(s) will reveal much about the regulation of the reproductive manipulations induced by Wolbachia.

New Protein Domain Predictions Are Variable across the Cif Phylogeny

We recovered the four previously identified phylogenetic Types (LePage et al. 2017). Here, our analyses include additional strains that cause reproductive parasitism beyond CI (parthenogenesis and male-killing, table 1), and the more divergent Type IV paralogs for cifA, so far identified in B-Supergroup Wolbachia. We recover a set of Type III alleles from wUni, a strain that induces parthenogenesis in the parasitoid wasp, Muscidifurax uniraptor (Stouthamer et al. 1993). The wBol1-b strain, a male-killer that has retained CI
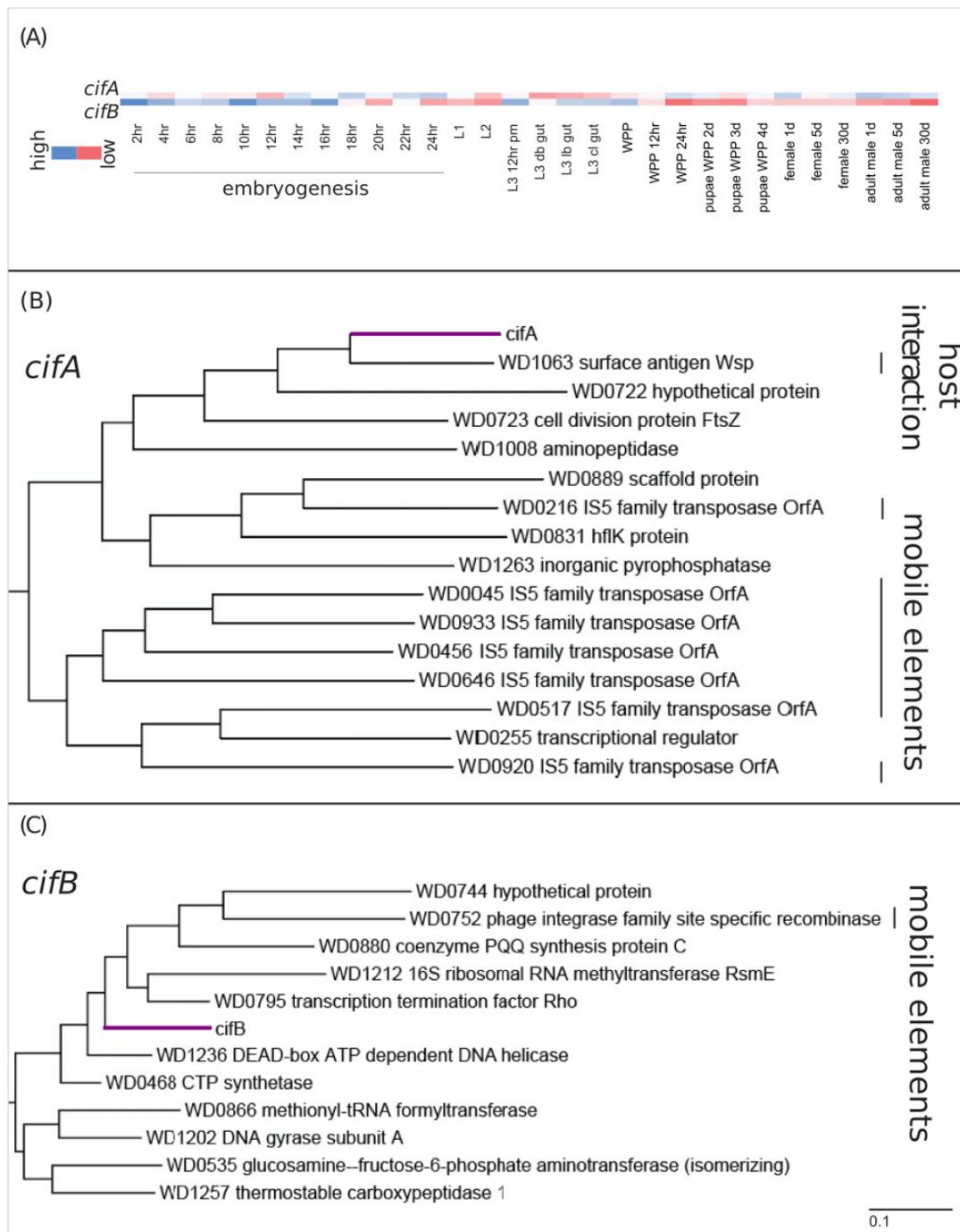
FIG. 3.—Gene expression of *cifA* and *cifB* during *Drosophila melanogaster* development. (A) Heatmap representation of normalized transcripts per base pair per million (TPM) for both *cifA* and *cifB* during *Drosophila melanogaster* development. *cifB* is highly expressed during embryogenesis and downregulated after pupation, whereas *cifA* is more highly expressed in adults and pupae. Clustering of *Wolbachia* loci based on expression across fly development illustrates correlated expression profiles between *w*Mel loci and *cifA* (B) or *cifB* (C). Mobile elements and loci involved in host interaction (*wsp*) are indicated with vertical lines on the right side of the figure.

capabilities (Hornett et al. 2008), has alleles belonging to both Type I and Type IV.

Homologs and predicted protein domains of CifA and CifB for all four phylogenetic Types (LePage et al. 2017) from

Wolbachia strains that cause CI, parthenogenesis, male-killing, or no reproductive phenotype were characterized by HHpred homology and domain structure prediction software (Söding et al. 2005). Search parameters are described in the

methods. Several new prominent protein domains, herein referred to as "modules," were identified for each CifA and CifB protein sequence (table 2).

For CifA, three modules were annotated (fig. 4A, table 2). First, the most N-terminal module (ModA-1) is only recovered in Type I variants, with distant homology (~22% amino acid identity) to Catalase-rel that is predicted to catalyze the breakdown of hydrogen peroxide (Chelikani et al. 2004). The probability of the module being homologous to Catalase-rel is low (prob = 21–24), but the consistent recovery of structure in this region across Type I alleles is notable. The second CifA module in the central region (ModA-2) has homology to a domain of unknown function (Types I, II, and IV, prob = 27–64), globin-like domains (Types II and IV, prob = 21–30), and Puf family RNA-binding domains (Types III and IV, prob = 25–49). The last CifA module in the C-terminal region (ModA-3) has hits to an STE-like transcription factor in all Types (prob = 27–42). In general, the CifA proteins showed distant homology to known domains, but we consistently recovered the same regions of structure within CifA protein Types.

For CifB, three modules were also defined (fig. 4B, table 2). The first (ModB-1) and second (ModB-2) most N-terminal regions of all Types both have matches to the PDDEXK nuclease family (prob = 57–98) and various other restriction endonucleases such as NucS, HSDR_N, and MmcB (prob = 50–91). The third module, found only in the Type I C-terminus (ModB-3), has very strong homology to a number of ubiquitin-modification and protease-like domains (prob = 71–96). This was expected, as ModB-3 contains the predicted catalytic residue associated with CI function in CidB, known to have deubiquitylase activity (Beckmann et al. 2017). WOVitA4 (Type 1) has an extended C-terminus not present in any other alleles, and within that extended C-terminus is an additional structural domain, with homology to a Herpesvirus tegument protein (prob = 53), and a phosphohydrolase-associated domain (prob = 57). CifB Type IV alleles (WOAlbB, WOPip2, and wBol1-b) were not included in the phylogenetic reconstruction, as they are highly divergent and not reciprocal blasts of WOMelB cifB. Despite their divergence, these Type IV CifB alleles have similar structures to Type II and III alleles: Two PDDEXK-like modules, and no Ulp-1-like module 3 (supplementary fig. S3, Supplementary Material online). Full structural schematics with exact coordinates and homology regions for each allele are available in the Supplementary Material online (supplementary figs. S2 and S3, Supplementary Material online), as are all significant domain hits with associated probabilities and extended descriptions (supplementary tables S1 and S2, Supplementary Material online).

CifA and CifB Codiverge

Initial phylogenetic trees based on core amino acid sequences of Type I–III variants of CifA and CifB exhibited similar trees

Table 2

Predicted Structural Modules of Cif Proteins

Protein	Module[a]	Size Range (AA)	Homology
CifA	ModA-1[b] ⬤	21–22	• Catalase-rel, decomposes hydrogen peroxide into water and oxygen
	ModA-2 ⬤	65–264	• DUF3243 domain of unknown function • Puf family RNA binding • Globin-like protein
	ModA-3 ⬤	47–74	• STE-like transcription factor
CifB	ModB-1 ⬤	103–133	• PDDEXK, PD-(D/E)XK nuclease superfamily • Endonuclease NucS • Restriction endonuclease-like family • HSDR_N, type I restriction enzyme R protein N terminus
	ModB-2 ⬤	122–205	• PDDEXK, PD-(D/E)XK nuclease superfamily • MmcB-like DNA repair protein • COG5321, uncharacterized protein • HSDR_N, type I restriction enzyme R protein N-terminus • Endonuclease NucS
	ModB-3[b] ○	95–147	• Ulp-1, ubiquitin-like proteases • Various proteases and peptidases (C5, C57, Sentrin-specific protease)

[a]Colors next to modules are used throughout the Figures.
[b]Only present in Type I.

(LePage et al. 2017). Here, we statistically ground the inference of codivergence using the largest set of Wolbachia homologs to date. We quantified congruence between the CifA and CifB phylogenetic trees for Types I–III (supplementary file S1, Supplementary Material online) using MC and RF tree metrics (Robinson and Foulds 1981; Bogdanowicz et al. 2012; Bogdanowicz and Giaro 2013), with normalized distances ranging from 0.0 (complete congruence) to 1.0 (complete incongruence). Results show strong levels of congruence between CifA and CifB ($P < 0.00001$ for both, normalized MC = 0.06 and normalized RF = 0.125). To further statistically validate the inference of codivergence, we measured the correlation between patristic distance matrices for CifA and CifB using the Mantel test (Mantel 1967). Results demonstrate a high degree of correlation between patristic distance matrices, and through permutation show that independent evolution of CifA and CifB is highly unlikely (Pearson correlation coefficient = 0.905, $P = 0.00001$).
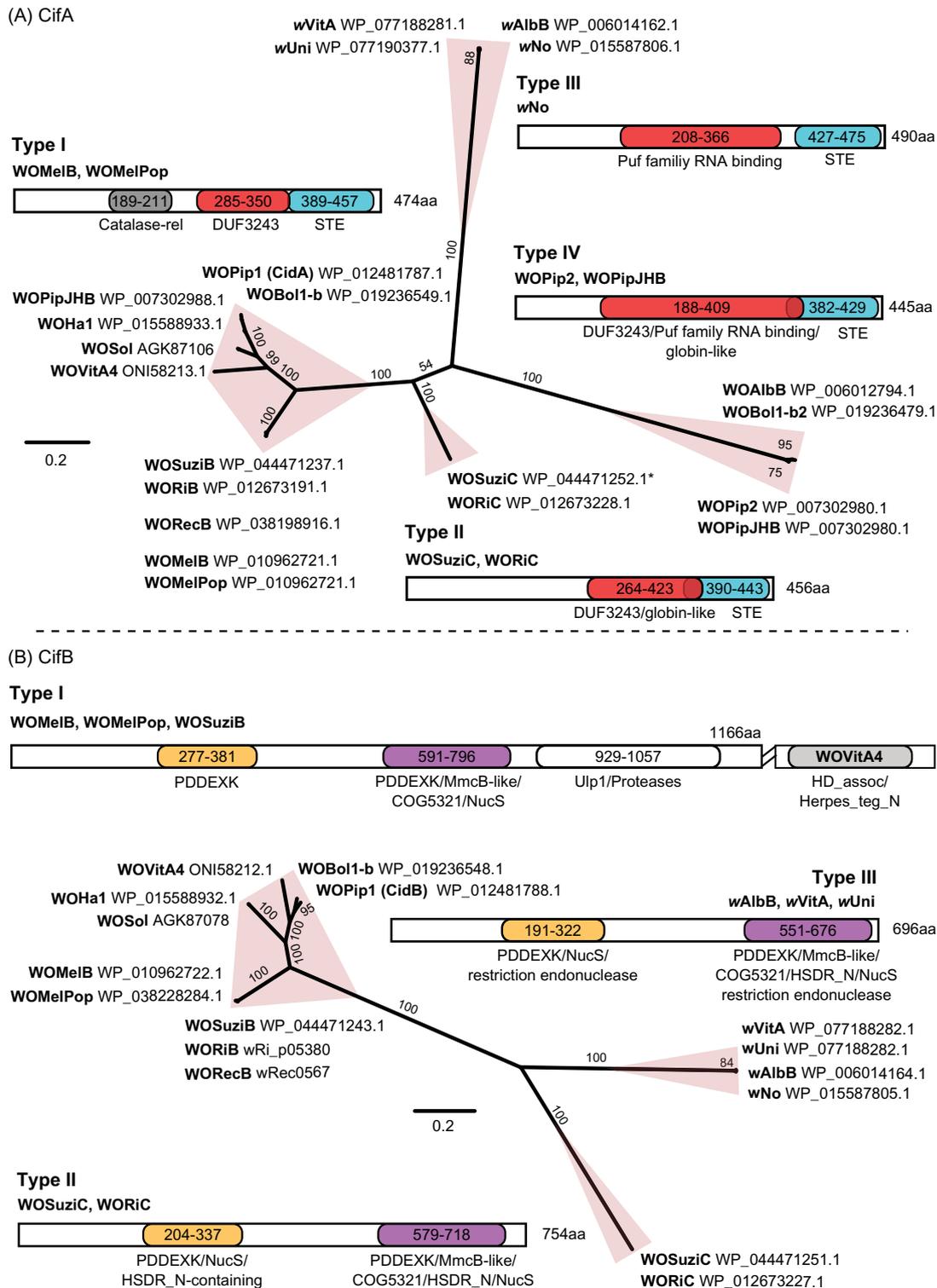
FIG. 4—Phylogenetic relationships and representative predicted protein structure of Cif protein Types. (A) CifA and (B) CifB. Alleles are in bold next to their corresponding accession number, and pink shapes around branches designate monophyletic "Types." Representative structures are shown for each type, with the length of the protein indicated at the C-terminus. If genes differed by only a few amino acids a single representative is shown. Allele names use the previously described naming convention with a WO prefix referring to particular phage WO haplotype, and the *w* prefix indicating a phage WO-like island (LePage et al. 2017). The N-terminus of CifA WOSuziC (* in figure) was translated from the end of another contig and concatenated to get the full-length protein (see Materials and Methods). WOMelB and WOMelPop are identical at the amino acid level, as are WOPipJHB and WOPip2.
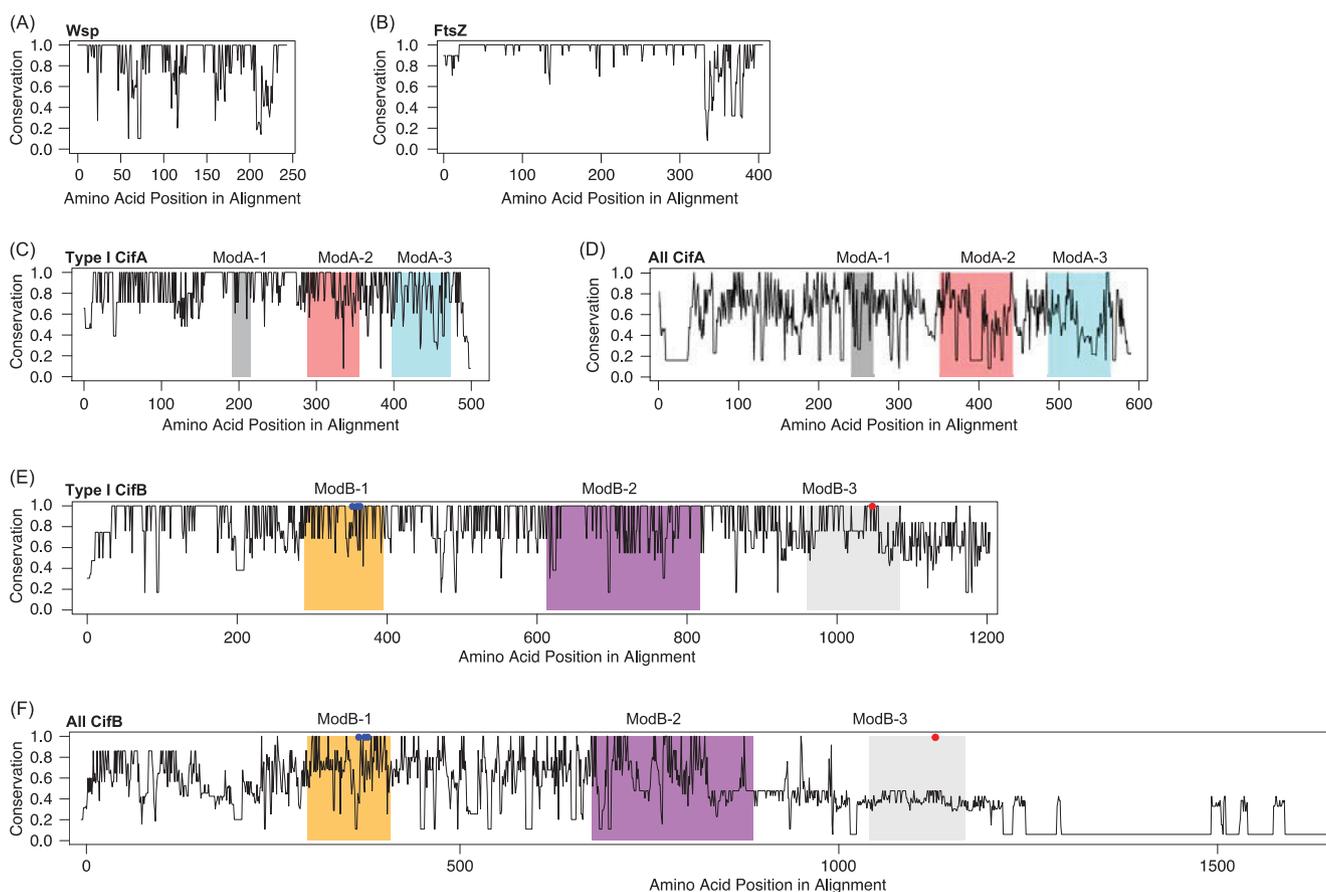
Fig. 5—Protein conservation, as determined by Shannon entropy scores. (A) Wsp, (B) Cell division protein FtsZ, (C) Type I CifA, (D) All CifA, (E) Type I CifB alleles except for WOVitA4, (F) All CifB alleles. Red dots in (E) and (F) indicate the ModB-3 catalytic residue (Beckmann et al. 2017), unique to and completely conserved for Type I alleles. Blue dots in (E) and (F) represent the (P)D-(D/E)XK motif (Kosinski et al. 2005) present in wMel.

Cif Proteins Evolve Rapidly

Amino acid sequence conservation across the full length of the Cif proteins was determined and compared with Wolbachia amino acid sequences of genes that either have signatures of recombination and directional section (Wsp) or have not undergone extensive recombination and directional selection (FtsZ, cell division protein). Wsp protein sequences exhibit considerable divergence (mean conservation = 0.85), with very few sites in a row being completely conserved (fig. 5A). In contrast, FtsZ is relatively conserved (mean conservation = 0.94), and most of the divergence is clustered at the C-terminus (fig. 5B). Mean conservation for the Cif protein sequences was lower than Wsp—0.83 for Type I CifA alleles (fig. 5C) and 0.82 for Type I CifB alleles (fig. 5E, table 3). When all Cif alleles were considered, mean conservation was even further reduced—0.58 for CifA (fig. 5D) and 0.43 for CifB (fig. 5F). The lower average conservation of CifB genes is in part due to the many insertions and deletions in the alignment, and the missing C-terminal deubiquitylase region, ModB-3, of the Type II and III alleles. Thus, several CifB proteins apparently lack this activity, and whether these variants

Table 3
Average Amino Acid Conservation of Cifs and Modules

Protein	Region[a]	Type I	All
CifA	ModA-1[b]	0.93	0.67
	ModA-2	0.84	0.53
	ModA-3	0.77	0.53
	CifA	0.83	0.58
CifB	ModB-1	0.89	0.70
	ModB-2	0.87	0.60
	ModB-3[b]	0.79	0.40
	CifB	0.82	0.43

[a]Module number is defined in table 2.
[b]Only annotated in Type I.

cause CI remains to be determined. Although the CifB proteins are highly divergent, the catalytic residue (red dot in fig. 5E and F) in the deubiquitylating module of CifB is unique to and completely conserved for the Type I alleles.

The Cif proteins have extensive amounts of diversity, with completely conserved amino acids distributed across the length of the protein, and not confined to any particular

regions (fig. 5C–F, supplementary tables S3–S6, Supplementary Material online). There were significant differences in the level of conservation between modules and non-module regions for the Type I alignments of both CifA ($F_{3,\ 490}$ = 4.276, $P = 0.0054$) and CifB ($F_{3,\ 1195}$ = 9.703 $P = 1.5e-06$) (table 3). The only modules that had significantly higher conservation than the nonmodule regions of the alignment were ModB-1 ($P = 0.0173$) and ModB-2 ($P = 0.0011$). The *w*Mel strain contains the (P)D-(D/E)XK motif in ModB-1 (blue dots in fig. 5E and F) (Kosinski et al. 2005), but it is less than 80% conserved across strains despite the higher average conservation of this module. In contrast, *w*Mel does not contain the catalytic motif in ModB-2, also a PDDEXK nuclease-like domain. The ModB-2 (P)D-(D/E)XK motif is present in Type IV alleles such as WOPip2 that when mutated no longer induces growth defects in yeast (Beckmann et al. 2017). ModA-3 is significantly less conserved than the nonmodule regions of Type I CifA ($P = 0.0300$).

Cif Module Presence Generally Predicts Reproductive Phenotype

We used the *w*Mel-predicted Cif modules as a seed to search for the presence of homologous modules across *Wolbachia* genome sequences using TBlastN (fig. 6), with the intent of discovering *cif*-like regions or remnants in strains with other phenotypes. In strains with more divergent Cif Types, we report modules that were expected based on the HHpred results, but not recovered with TBlastN due to sequence divergence from WOMelB. Additionally, we recover homologous modules outside of the annotated *cif* open reading frames, such as the chromosomal region with a ModB-3 (Ulp-1-like) region in *w*No. The ModB-3 *w*No module is genic, found within a hypothetical protein (WP_041581315.1). Whether or not these *cif*-like regions outside of prophage WO contribute to CI remains to be determined. All Supergroup A and B strains, with the exception of *w*Au and *w*Tpre (non-CI inducing strains), contained at least one recovered module.

Importantly, all strains that are known to be capable of inducing or rescuing CI have two or more recovered modules, though they do not necessarily have ModB-3, which contains the catalytic residue implicated in CI function (Beckmann et al. 2017). The non-CI strains have fewer recovered modules: One module (ModB-2) in *w*Uni, and no modules in *w*Cle, *w*Au, *w*Tpre and the nematode-infecting strains (Supergroups C and D). *w*Uni is a unique case, where we identified *cif* alleles in the genome, but recovered only one module. Most *w*Uni modules are either missing (fig. 4A) or divergent enough from WOMelB that they were not considered a positive match. *w*AlbB and *w*No, both CI-inducing strains with Type III and IV alleles, have fewer recovered modules, but this is congruent with the more divergent nature of those Cif Types. It is notable that despite the phylogenetic distance from WOMelB,

more modules are recovered from the CI-inducing strains than the non-CI inducing strains within the same Type.

The high number of modules in *w*Suzi and *w*Ri is due to the presence of a duplicated set of Type I variants. We recovered many modules in *w*Suzi, which is a strain not known to induce CI, but is sister to *w*Ri, which can induce CI (Hamm et al. 2014; Cattel et al. 2016). This discrepancy between *cif* presence and absence of a reproductive phenotype might be explained by the disrupted Type II *cifA* in *w*Suzi. The split WOSuziC sequenced was concatenated to allow for a more robust phylogenetic reconstruction (fig. 4), but it is in fact disrupted by a transposase (Conner et al. 2017). However, having a functional set of Type I *cif* alleles appears to be sufficient for CI-induction in other strains (Beckmann et al. 2017; LePage et al. 2017), so it is not clear how inactivation of the Type II alleles here may affect the final CI phenotype in *w*Suzi. Strain *w*Di, infecting the Asian citrus psyllid *Diaphorina citri*, has no identified reproductive phenotype, but only contains two modules: ModB-1 and ModB-3. For *w*Ha, we recovered duplicates of all the modules. These represent a highly disrupted copy of the gene set harboring frameshifts that were annotated as pseudogenes.

The lack of evidence for homologous *cif* genes in the C, D, and F Supergroup *Wolbachia* agrees with previous findings (LePage et al. 2017) that CI-function is restricted to the A + B-Supergroup clade (likely due to WO phage activity), and the absence of WO phages for the nematode-infecting strains (Gavotte et al. 2007). The loss of CI within the A and B Supergroups is likely a derived trait due to the rapid evolution of prophage WO (Ishmael et al. 2009; Kent, Salichos, et al. 2011) and relaxed selection after transition to a new reproductive phenotype. The low number of modules identified in such strains is consistent with gene degradation and loss. Additionally, we recover no *cif*-like regions in *Cardinium*, a member of the *Bacteroidetes*, and an independent transition to a CI phenotype (Penz et al. 2012).

To further explore the conservation of the *cif* genes across the sequenced *Wolbachia*, and to uncover diversity that may be present in other genomes, we searched the WGS databases for recently sequenced genomic scaffolds from *Wolbachia* infecting the *Nomada* bees (*w*Nleu, *w*Nla, *w*Npa, and *w*Nfe) (Gerth and Bleidorn 2016), *Drosophila incompta* (*w*Inc_Cu) (Wallau et al. 2016), and *Laodelphax striatellus* (*w*Stri) (GenBank accession number NZ_LRUH00000000.1) using HMMER. Only for *w*Stri do we have direct evidence of CI induction (Noda et al. 2001) yet the *w*Stri WGS projects contain only one *cif* locus (CifB) with an unusual structure not found in any of the other Types. On the basis of HHpred analyses, the *w*Stri homolog contains a deubiquitylase region in the middle of the protein, with two downstream regions that have homology to glucosyl transferases and lipases, respectively (supplementary fig. S4, Supplementary Material online). The *w*Inc_Cu WGS project contained one each of CifA and CifB alleles. The CifA allele from *w*Inc_Cu is a typical Type
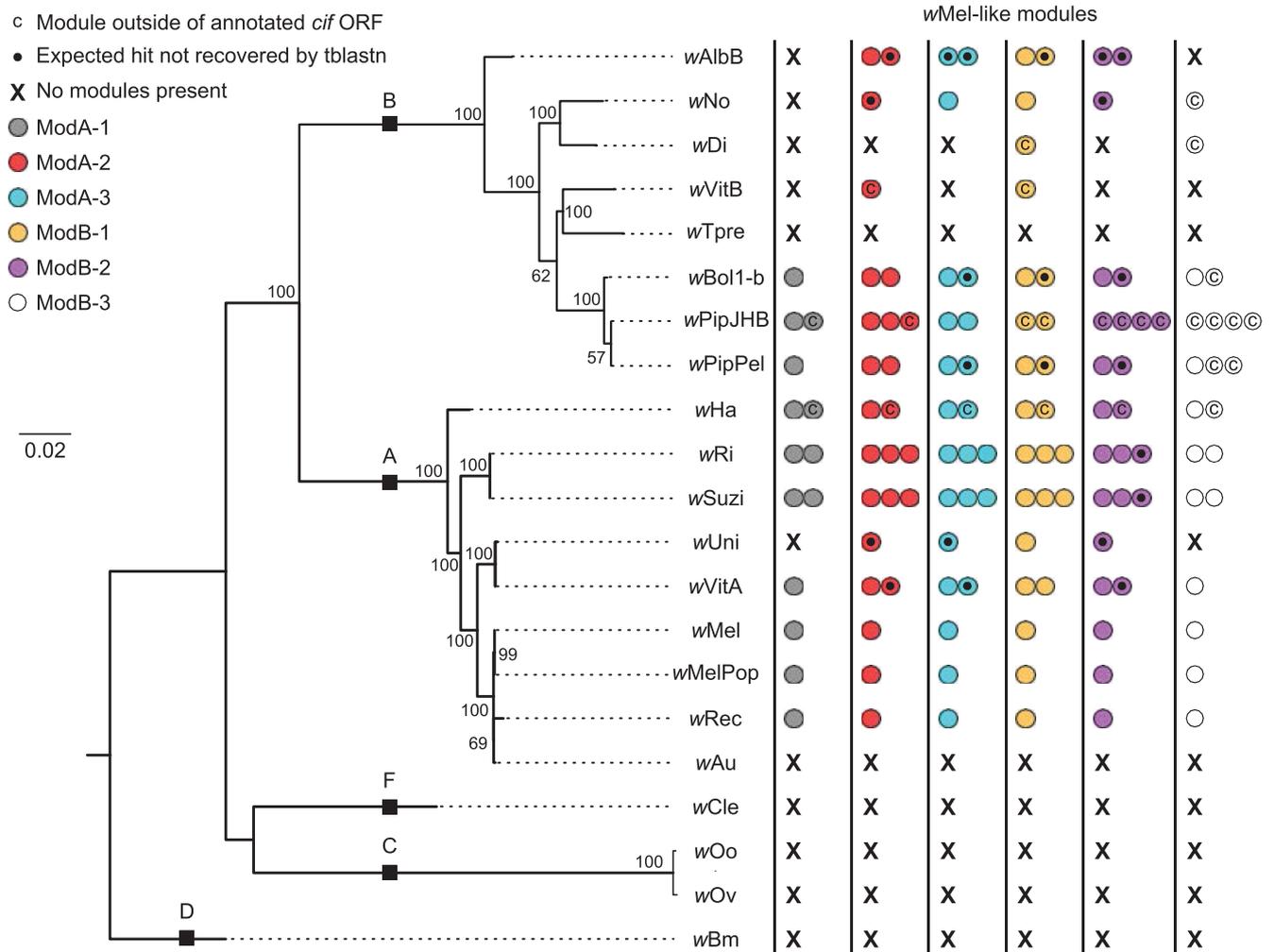
FIG. 6—Presence of wMel-like Cif modules across the Wolbachia phylogeny. The WOMelB module sequences were used to query available Wolbachia genomes to look for the presence of Cif-like regions beyond those within the annotated Cifs (fig. 5). Colored dots correspond to the structural regions delimited by HHpred, shown in figure 4, and listed in table 2. A "C" within a dot indicates the presence of a module outside of annotated cif open reading frames (fig. 4 and supplementary figs. S2 and S3, Supplementary Material online). The black dot indicates a module annotated by HHpred, but not identified by TBlastN due to divergence from the WOMelB module. Black boxes labeled with uppercase letters indicate branches leading to Wolbachia Supergroups. Dotted lines on the phylogeny lead to taxon names and are not included in the branch length.

I protein containing three modules: An N-terminal Catalase-rel domain and an internal DUF3242 domain, followed by the STE-like transcriptional factor domain. Because these are incomplete genome projects, it is possible that other cif homologs have been missed due to the current sequencing coverage. Alternatively, it is possible that other, as yet undiscovered, mechanisms of reproductive manipulation exist in these strains. In contrast, the Nomada-associated Wolbachia contain a large repertoire of cif homologs, including Types I, II, IV, and several homologs with variations on the Type IV domain architecture for CifA (supplementary fig. S4, Supplementary Material online). Many of the CifB homologs are disrupted Type I variants that contain the deubiquitylase-like domain, but not the nuclease-like domains (supplementary fig. S4, Supplementary Material online).

Discussion

We explored three key features of cif evolution: 1) The operon hypothesis, 2) potential novel functions across the cifA and cifB phylogenies, and 3) the conservation and diversity of cif alleles across strains with different host-manipulation phenotypes. We provide multiple lines of evidence that although cifA and cifB can be cotranscribed, they are divergently transcribed in wMel during host development, suggesting a more complex regulation of gene expression than found in classical operons. Indeed, cifB transcription has a significant negative correlation with cifA. In Escherichia coli, operons frequently have internal promoters and terminators that result in different units of transcription, which are preferentially used during certain conditions (Conway et al. 2014). Although cifB is

expressed at about 1/10 the level of *cifA* across all life cycle stages, the significant negative correlation in their levels suggests that the same factor(s) could upregulate *cifA* while downregulating *cifB*, or a *cifA* encoded RNA or protein may inhibit *cifB* expression or vice versa. The new annotations for *cifA* alleles, including a Puf family RNA-binding-like domain and STE transcription factor, could theoretically play a role in inhibiting *cifB* expression. More detailed analyses from a variety of strains, *cif* Types, and conditions would help develop a comprehensive understanding of the factors regulating expression of these genes and the CI phenotype.

It is especially interesting that *cifA* and *cifB* synteny is maintained across prophage WO regions, despite the high level of recombination and rearrangements in prophage WO and *Wolbachia* genomes (Baldo, Bordenstein, et al. 2006; Kent, Funkhouser, et al. 2011; Ellegaard et al. 2013). Although it is not yet clear why *cifA* and *cifB* homologs maintain their syntenic orientation, given that they have very different absolute and relative expression levels, we hypothesize that this feature can be attributed to 1) their location within prophage WO and/or 2) functions associated with the ability of *cifA* and *cifB* to act synergistically to induce CI (LePage et al. 2017), or 3) with the potential antagonism of *cifA* on *cifB* transcripts or transcription. For example, they could share the same promoter, but in addition to the Rho-independent terminator, *cifA* may further inhibit *cifB* expression by binding to the intergenic region between them, causing the polymerase to terminate. Such a model would be consistent with both the absolute expression differences and the negative correlation. Alternatively, *cifA* and *cifB* may have different promoters, but a bicistronic message occurs because of an imperfect hairpin terminator between the genes. We conclude that *cifA* and *cifB* are cotranscribed but not coregulated as in a classical operon, and do not act strictly as a toxin–antitoxin system due to the requirement of both Cif proteins for the induction of CI in arthropods. Determining how *cifA* and *cifB* expression is regulated in the insect host will advance an understanding of both the basic biology of CI and vector control programs that deploy CI to control disease transmission.

Despite the conservation of gene order, Cif proteins showed extensive amounts of divergence and differences in domain structure as previously reported (LePage et al. 2017). Here, the levels of amino acid conservation in the Cifs are lower than FtsZ and Wsp, the latter of which is known to recombine and be subject to directional selection. The conservation of the predicted catalytic residue in the C-terminal deubiquitylase domain is an important feature of CidB (Beckmann et al. 2017). Although this residue is required for CI induction in CidB and other Type I alleles (Beckmann et al. 2017), only Type I alleles (of the four identified Types) have this domain. Strains known to induce CI, such as *w*AlbB and *w*No do not have Type I alleles, implying that the deubiquitylase domain is not essential for inducing CI across other Cif Types. The complete, functional capacity of all Types has

yet to be explored *in vivo*, but is a promising direction for understanding the evolution of *Wolbachia*–host associations.

Based on what is known about *Wolbachia* biology, some of the protein domains may be especially good candidates for further study and *in vivo* functional characterization. Predicted PDDEXK-like nuclease domains are present in all four CifB Types. Given the predicted interaction of these domains with DNA (Kosinski et al. 2005), and the presence of these domains across CifB proteins, determining whether and how these regions interact with host (*Wolbachia* or insect) DNA, and whether or not they contribute to CI function would be useful in understanding the consistent presence of this module. Mutating the predicted catalytic site of the nuclease region in *w*Pip's Type IV CifB (aka CinB) reduces toxicity in yeast (Beckmann et al. 2017). However, this catalytic residue is not conserved, so further exploration of nuclease function across more divergent alleles will be useful. As aforementioned, many of the CifA alleles encode Puf family RNA-binding-like domains, which have previously been implicated in mRNA localization and transcriptional regulation (Quenault et al. 2011). This RNA binding-like domain is found upstream of an STE transcription factor-like domain and could provide a promising direction for understanding the complicated transcriptional dynamics of the *cif* genes.

Wolbachia strains that have lost CI have a strong signature of *cif* gene degradation and loss, consistent with their role in CI. The two parthenogenesis-inducing strains (*w*Tpre and *w*Uni) appear to be at different places in this process of gene loss, with divergent Cif amino acid sequences recovered for *w*Uni, but no modules identified in *w*Tpre. There are several explanations for this. *w*Uni is likely a more recent transition to parthenogenesis, as it is closely related to a CI strain (*w*VitA) (Baldo, Dunning Hotopp, et al. 2006; Newton et al. 2016). In comparison, *w*Tpre is part of a unique clade of *Wolbachia* that all induce parthenogenesis in *Trichogramma* wasps (Rousset et al. 1992; Werren et al. 1995; Schilthuizen and Stouthamer 1997). This strain has lost its WO phage association and only has relics of WO phage genes (Gavotte et al. 2007; Lindsey et al. 2016). Additionally, the two strains that independently transitioned to the parthenogenesis phenotype have evolved separate mechanisms for doing so (Stouthamer and Kazmer 1994; Gottlieb et al. 2002). Differences in time since transition to the parthenogenesis phenotype, phage WO associations, and mechanisms of parthenogenesis induction likely all play a role in the rate of *cif* gene degradation.

Although *cifA* and *cifB* are prophage WO genes, not all CI-inducing strains have a complete prophage. Indeed, the *w*Rec strain of *Wolbachia* in *Drosophila recens* is one such example where approximately three quarters of prophage WO genes were eliminated (Metcalf et al. 2014), previously resulting in failed detection of WO presence (Bordenstein and Wernegreen 2004). However, genomic analyses of phage WO particles from wasps and moths revealed that several

genes packed in the genome of phage WO particles (Bordenstein SR and Bordenstein SR 2016) are in fact retained in prophage WO of wRec (Metcalf et al. 2014), including *cifA* and *cifB*. Genes in prophage WO relics are apparently a source of host-manipulation across *Wolbachia* genomes. Additionally, there is considerable variation in the strength of CI across different *Wolbachia* strains (Veneti et al. 2003). CifA and CifB have an additive effect on the strength of CI (LePage et al. 2017), so it is possible that the level of *cif* expression, or the ratio of *cifA* and *cifB* transcripts across development, are ways in which CI strength is adjusted. The rapidly evolving nature of the Cif proteins may affect other ways in which they function in the host. For example, in Type I CifB proteins that have the essential deubiquitylase residue, other sequence variation may affect the ability to bind with CifA, locations of posttranslational modifications, or the ability to be efficiently localized to the host nucleus. Additionally, the level of CI is often host-dependent (Bordenstein and Werren 1998; McGraw et al. 2001), possibly a result of how well *Wolbachia* replicate in the host, and/or the specificity of Cif proteins with the host target, which is currently unknown. There are also environmental conditions that affect the strength of CI, and they likely do so by affecting *Wolbachia* titers and resulting Cif expression in the host (Clancy and Hoffmann 1998; Yamada et al. 2007).

On the basis of our analyses, we propose three avenues of research on the function of the Cif proteins. First, functional confirmation of the newly annotated modules will be important in understanding how these genes function enzymatically. In total, we predict six modules in the Cif protein sequence homologs, with varying degrees of confidence (supplementary tables S1 and S2, Supplementary Material online). For some of these modules, straightforward experiments can be designed in model systems (such as *Saccharomyces*) to determine whether their predicted function is correct, as has been done for the deubiquitylase domain of CidB (Beckmann et al. 2017) and countless other bacterial effectors (Kramer et al. 2007; Siggers and Lesser 2008; Archuleta 2011). The necessity and importance of these modules to the CI phenotype can be assessed in the *Drosophila* model, where the induction of the phenotype and rescue is straightforward (LePage et al. 2017). Second, detailed characterization of *cif* gene regulation will be important for understanding CI expression and penetrance, thus informing vector control programs that rely on proper expression of the CI phenotype, often in a transfected host. Finally, we suggest that although the discovery of these genes is fundamental, it is clear from this analysis that we have not comprehensively evaluated or identified the mechanisms behind CI and other reproductive manipulations. The gene characterization analyses described here reveal new and relevant annotations, but with many regions of unknown function across all of the phylogenetic Types, missing deubiquitylase domains in particular CI strains, and a coevolving, phylogenetic relationship across the Cif

trees. Importantly, the locus, presumably expressed in the female insect infected with a compatible *Wolbachia*, and mechanism behind rescuing CI are still unknown, as is the exact mechanism by which all Cif proteins induce CI. Therefore, the recent discovery of these CI genes and their sequence characterization described here pave the way for investigating key mechanisms of the *Wolbachia*–host symbiosis.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Literature Cited

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21(9):2104–2105.

Archuleta TL. 2011. The *Chlamydia* effector chlamydial outer protein N (CopN) sequesters tubulin and prevents microtubule assembly. J Biol Chem. 286(39):33992–33998.

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

Baldo L, Dunning Hotopp JC, et al. 2006. Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. Appl Environ Microbiol. 72(11):7098–7110.

Baldo L, Lo N, Werren JH. 2005. Mosaic nature of the *Wolbachia* surface protein. J Bacteriol. 187(15):5406–5418.

Bandi C, Anderson TJC, Genchi C, Blaxter ML. 1998. Phylogeny of *Wolbachia* in filarial nematodes. Proc R Soc Lond B. 265(1413):2407–2413.

Beckmann JF, Fallon AM. 2013. Detection of the *Wolbachia* protein WPIP0282 in mosquito spermathecae: implications for cytoplasmic incompatibility. Insect Biochem Mol Biol. 43(9):867–878.

Beckmann JF, Ronau JA, Hochstrasser M. 2017. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. Nat Microbiol. 2:17007.

Bian G, et al. 2010. The endosymbiotic bacterium *Wolbachia* induces resistance to dengue virus in *Aedes aegypti*. PLoS Pathog. 6(4):e1000833.

Bogdanowicz D, Giaro K. 2013. On a matching distance between rooted phylogenetic trees. Int J Appl Math Comp Sci. 23:669–684.

Bogdanowicz D, Giaro K, Wróbel B. 2012. TreeCmp: comparison of trees in polynomial time. Evol Bioinform Online. 8:EBO.S9657.

Bordenstein SR, Bordenstein SR. 2016. Eukaryotic association module in phage WO genomes from *Wolbachia*. Nat Commun. 7. doi:10.1038/ncomms13155.

Bordenstein SR, Marshall ML, Fry AJ, Kim U, Wernegreen JJ. 2006. The tripartite associations between bacteriophage, *Wolbachia*, and arthropods. PLoS Pathog. 2(5):e43.

Bordenstein SR, O'Hara FP, Werren JH. 2001. *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. Nature 409(6821):707–710.

Bordenstein SR, Wernegreen JJ. 2004. Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. Mol Biol Evol. 21(10):1981–1991.

Bordenstein SR, Werren JH. 1998. Effects of A and B *Wolbachia* and host genotype on interspecies cytoplasmic incompatibility in *Nasonia*. Genetics 148:1833–1844.

Bourtzis K, et al. 2014. Harnessing mosquito–*Wolbachia* symbiosis for vector and disease control. Acta Trop. 132:S150–S163.

Brooks AW, Kohl KD, Brucker RM, van Opstal EJ, Bordenstein SR. 2016. Phylosymbiosis: relationships and functional effects of microbial communities across host evolutionary history. PLoS Biol. 14(11):e2000225.

Brucker RM, Bordenstein SR. 2013. The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus *Nasonia*. Science 341(6146):667–669.

Caporaso JG, et al. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 7(5):335.

Capra JA, Singh M. 2007. Predicting functionally important residues from sequence conservation. Bioinformatics 23(15):1875–1882.

Cattel J, et al. 2016. *Wolbachia* in European populations of the invasive pest *Drosophila suzukii*: regional variation in infection frequencies. PLoS One 11(1):e0147766.

Chafee ME, Funk DJ, Harrison RG, Bordenstein SR. 2010. Lateral phage transfer in obligate intracellular bacteria (*Wolbachia*): verification from natural populations. Mol Biol Evol. 27(3):501–505.

Chelikani P, Fita I, Loewen PC. 2004. Diversity of structures and properties among catalases. Cell Mol Life Sci. 61(2):192–208.

Clancy DJ, Hoffmann AA. 1998. Environmental effects on cytoplasmic incompatibility and bacterial load in *Wolbachia*-infected *Drosophila simulans*. Entomol Exp Appl. 86(1):13–24.

Conner WR, et al. 2017. Genome comparisons indicate recent transfer of *w*Ri-like *Wolbachia* between sister species *Drosophila suzukii* and *D. subpulchrella*. Ecol Evol 7(22):9391.

Conway T, et al. 2014. Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. MBio 5(4):e01442–e01414.

Dedeine F, et al. 2001. Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. Proc Natl Acad Sci U S A. 98(11):6247–6252.

Duron O, Fort P, Weill M. 2006. Hypervariable prophage WO sequences describe an unexpected high number of *Wolbachia* variants in the mosquito *Culex pipiens*. Proc R Soc Lond B. 273(1585):495–502.

Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol. 7(10):e1002195.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.

Ellegaard KM, Klasson L, Näslund K, Bourtzis K, Andersson SGE. 2013. Comparative genomics of *Wolbachia* and the bacterial species concept. PLoS Genet 9(4):e1003381.

Felsenstein J. 1989. PHYLIP-phylogeny inference package (version 3.2). Cladistics 5:6.

Gautheret D, Lambert A. 2001. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. J Mol Biol. 313(5):1003–1011.

Gavotte L, et al. 2007. A survey of the bacteriophage WO in the endosymbiotic bacteria *Wolbachia*. Mol Biol Evol. 24(2):427–435.

Gerth M, Bleidorn C. 2016. Comparative genomics provides a timeframe for *Wolbachia* evolution and exposes a recent biotin synthesis operon transfer. Nat Microbiol. 2:16241.

Gottlieb Y, Zchori-Fein E, Werren JH, Karr TL. 2002. Diploidy restoration in *Wolbachia*-infected *Muscidifurax uniraptor* (Hymenoptera: pteromalidae). J Invertebr Pathol. 81(3):166–174.

Gutzwiller F, et al. 2015. Dynamics of *Wolbachia pipientis* gene expression across the *Drosophila melanogaster* life cycle. G3 5:2843–2856.

Hamm CA, et al. 2014. *Wolbachia* do not live by reproductive manipulation alone: infection polymorphism in *Drosophila suzukii* and *D. subpulchrella*. Mol Ecol. 23(19):4871–4885.

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(2):215–220.

Hoerauf A, et al. 1999. Tetracycline therapy targets intracellular bacteria in the filarial nematode *Litomosoides sigmodontis* and results in filarial infertility. J Clin Invest. 103(1):11–18.

Hoffmann AA, et al. 2011. Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. Nature 476(7361):454–U107.

Hornett EA, et al. 2008. You can't keep a good parasite down: evolution of a male-killer suppressor uncovers cytoplasmic incompatibility. Evolution 62(5):1258–1263.

Hosokawa T, Koga R, Kikuchi Y, Meng XY, Fukatsu T. 2010. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. Proc Natl Acad Sci U S A. 107(2):769–774.

http://python.org. Python Language Reference, version 2.7.

Hughes GL, et al. 2011. *Wolbachia* infections are virulent and inhibit the human malaria parasite *Plasmodium falciparum* in *Anopheles gambiae*. PLoS Pathog. 7(5):e1002043.

Hurst GD, et al. 1999. Male–killing *Wolbachia* in two species of insect. Proc R Soc Lond B. 266(1420):735–740.

Hurvich CM, Tsai CL. 1993. A corrected Akaike information criterion for vector autoregressive model selection. J Time Ser Anal. 14(3):271–279.

Ishmael N, et al. 2009. Extensive genomic diversity of closely related *Wolbachia* strains. Microbiology 155(Pt 7):2211–2222.

Jaenike J, Dyer KA, Cornish C, Minhas MS. 2006. Asymmetrical reinforcement and *Wolbachia* infection in *Drosophila*. PLoS Biol 4(10):e325.

Kambris Z, Cook PE, Phuc HK, Sinkins SP. 2009. Immune activation by life-shortening *Wolbachia* and reduced filarial competence in mosquitoes. Science 326(5949):134–136.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Kearse M, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649.

Kent BN, Funkhouser LJ, Setia S, Bordenstein SR. 2011. Evolutionary genomics of a temperate bacteriophage in an obligate intracellular bacteria (*Wolbachia*). PLoS One 6:e24984.

Kent BN, Salichos L, et al. 2011. Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*) determined by targeted genome capture. Genome Biol Evol. 3(0):209–218.

Kosinski J, Feder M, Bujnicki JM. 2005. The PD-(D/E) XK superfamily revisited: identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. BMC Bioinformatics 6:172.

Kramer RW, et al. 2007. Yeast functional genomic screens lead to identification of a role for a bacterial effector in innate immunity regulation. PLoS Pathog. 3(2):e21.

Landmann F, Orsi GA, Loppin B, Sullivan W, Schneider DS. 2009. *Wolbachia*-mediated cytoplasmic incompatibility is associated with impaired histone deposition in the male pronucleus. PLoS Pathog. 5(3):e1000343.

Lassy CW, Karr TL. 1996. Cytological analysis of fertilization and early embryonic development in incompatible crosses of *Drosophila simulans*. Mech Dev. 57(1):47–58.

LePage D, Bordenstein SR. 2013. *Wolbachia*: can we save lives with a great pandemic? Trends Parasitol. 29(8):385–393.

LePage DP, et al. 2017. Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. Nature 543(7644):243–247.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.

Lindsey ARI, Werren JH, Richards S, Stouthamer R. 2016. Comparative genomics of a parthenogenesis-inducing *Wolbachia* symbiont. G3; 6(7):2113–2123.

Lo N, Casiraghi M, Salati E, Bazzocchi C, Bandi C. 2002. How many *Wolbachia* supergroups exist? Mol Biol Evol. 19(3):341–346.

Macke TJ, et al. 2001. RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Res. 29(22):4724–4735.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27:209–220.

Masui S, Kamoda S, Sasaki T, Ishikawa H. 2000. Distribution and evolution of bacteriophage WO in *Wolbachia*, the endosymbiont causing sexual alterations in arthropods. J Mol Evol. 51(5):491–497.

McClure R, et al. 2013. Computational analysis of bacterial RNA-Seq data. Nucleic Acids Res. 41(14): e140, https://doi.org/10.1093/nar/gkt444.

McGraw E, Merritt D, Droller J, O'Neill S. 2001. *Wolbachia*-mediated sperm modification is dependent on the host genotype in *Drosophila*. Proc R Soc Lond B. 268(1485):2565–2570.

Metcalf JA, Jo M, Bordenstein SR, Jaenike J, Bordenstein SR. 2014. Recent genome reduction of *Wolbachia* in *Drosophila recens* targets phage WO and narrows candidates for reproductive parasitism. PeerJ 2:e529.

Miller WJ, Ehrman L, Schneider D. 2010. Infectious speciation revisited: impact of symbiont-depletion on female fitness and mating behavior of *Drosophila paulistorum*. PLoS Pathog. 6(12):e1001214.

Min K-T, Benzer S. 1997. *Wolbachia*, normally a symbiont of *Drosophila*, can be virulent, causing degeneration and early death. Proc Natl Acad Sci U S A. 94(20):10792–10796.

Moreau J, Bertin A, Caubet Y, Rigaud T. 2001. Sexual selection in an isopod with *Wolbachia*-induced sex reversal: males prefer real females. J Evol Biol. 14(3):388–394.

Moreira LA, et al. 2009. A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, chikungunya, and *Plasmodium*. Cell 139(7):1268–1278.

Newton IL, et al. 2016. Comparative genomics of two closely related *Wolbachia* with different reproductive effects on hosts. Genome Biol Evol. 8(5):1526–1542.

Noda H, Koizumi Y, Zhang Q, Deng K. 2001. Infection density of *Wolbachia* and incompatibility level in two planthopper species, *Laodelphax striatellus* and *Sogatella furcifera*. Insect Biochem Mol Biol. 31(6–7):727–737.

O'Neill SL, Giordano R, Colbert AME, Karr TL, Robertson HM. 1992. 16S ribosomal-RNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. Proc Natl Acad Sci U S A. 89:2699–2702.

Penz T, et al. 2012. Comparative genomics suggests an independent origin of cytoplasmic incompatibility in *Cardinium hertigii*. PLoS Genet. 8(10):e1003012.

Pérez F, Granger BE. 2007. IPython: a system for interactive scientific computing. Comput Sci Eng. 9, doi: 10.1109/MCSE.2007.53.

Quenault T, Lithgow T, Traven A. 2011. PUF proteins: repression, activation and mRNA localization. Trends Cell Biol. 21(2):104–112.

Randerson JP, Jiggins FM, Hurst LD. 2000. Male killing can select for male mate choice: a novel solution to the paradox of the lek. Proc R Soc Lond B. 267(1446):867–874.

Robinson DF, Foulds LR. 1981. Comparison of phylogenetic trees. Math Biosci. 53(1–2):131–147.

Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61(3):539–542.

Ros VID, Fleming VM, Feil EJ, Breeuwer JAJ. 2009. How diverse is the genus *Wolbachia*? Multiple-gene sequencing reveals a putatively new *Wolbachia* supergroup recovered from spider mites (Acari: tetranychidae). Appl Environ Microbiol. 75(4):1036–1043.

Rousset F, Bouchon D, Pintureau B, Juchault P, Solignac M. 1992. *Wolbachia* endosymbionts responsible for various alterations of sexuality in arthropods. Proc R Soc Lond B. 250(1328):91–98.

Schilthuizen M, Stouthamer R. 1997. Horizontal transmission of parthenogenesis-inducing microbes in *Trichogramma* wasps. Proc R Soc Lond B. 264(1380):361–366.

scikit-bio.org. scikit-bio: core bioinformatics data structures and algorithms in Python. caporasolab.us: Northern Arizona University.

Serbus LR, Casper-Lindley C, Landmann F, Sullivan W. 2008. The genetics and cell biology of *Wolbachia*-host interactions. Annu Rev Genet. 42:683–707.

Shropshire JD, Bordenstein SR. 2016. Speciation by symbiosis: the microbiome and behavior. MBio 7(2):e01785-15.

Siggers KA, Lesser CF. 2008. The yeast *Saccharomyces cerevisiae*: a versatile model system for the identification and characterization of bacterial virulence proteins. Cell Host Microbe 4(1):8–15.

Sinkins SP, et al. 2005. *Wolbachia* variability and host effects on crossing type in *Culex* mosquitoes. Nature 436(7048):257–260.

Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33:W244–W248.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313.

Stouthamer R, Breeuwer JAJ, Luck RF, Werren JH. 1993. Molecular-identification of microorganisms associated with parthenogenesis. Nature 361(6407):66–68.

Stouthamer R, Kazmer DJ. 1994. Cytogenetics of microbe-associated parthenogenesis and its consequences for gene flow in *Trichogramma* wasps. Heredity 73(3):317–327.

Stouthamer R, Luck R. 1993. Influence of microbe-associated parthenogenesis on the fecundity of *Trichogramma deion* and *T. pretiosum*. Entomol Exp Appl. 67(2):183–192.

Stouthamer R, Luck RF, Hamilton WD. 1990. Antibiotics cause parthenogenetic *Trichogramma* (Hymenoptera, Trichogrammatidae) to revert to sex. Proc Natl Acad Sci U S A. 87(7):2424–2427.

Sutton E, Harris S, Parkhill J, Sinkins S. 2014. Comparative genome analysis of *Wolbachia* strain wAu. BMC Genomics 15:928.

Teixeira L, Ferreira Á, Ashburner M, Keller L. 2008. The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. PLoS Biol. 6(12):e1000002.

Tram U, Sullivan W. 2002. Role of delayed nuclear envelope breakdown and mitosis in *Wolbachia*-induced cytoplasmic incompatibility. Science 296(5570):1124–1126.

Turelli M, Hoffmann AA. 1991. Rapid spread of an inherited incompatibility factor in California *Drosophila*. Nature 353(6343):440–442.

Veneti Z, et al. 2003. Cytoplasmic incompatibility and sperm cyst infection in different *Drosophila-Wolbachia* associations. Genetics 164(2):545–552.

Walker T, et al. 2011. The wMel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. Nature 476(7361):450–U101.

Wallau GL, da Rosa MT, De Re FC, Loreto EL. 2016. *Wolbachia* from *Drosophila incompta*: just a hitchhiker shared by *Drosophila* in the New and Old World? Insect Mol Biol. 25(4):487–499.

Werren JH, Baldo L, Clark ME. 2008. *Wolbachia*: master manipulators of invertebrate biology. Nat Rev Microbiol. 6(10):741–751.

Werren JH, Windsor DM. 2000. *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? Proc R Soc Lond B. 267(1450):1277–1285.

Werren JH, Zhang W, Guo LR. 1995. Evolution and phylogeny of *Wolbachia*—reproductive parasites of arthropods. Proc R Soc Lond B. 261(1360):55–63.

Yamada R, Floate KD, Riegler M, O'Neill SL. 2007. Male development time influences the strength of *Wolbachia*-induced cytoplasmic incompatibility expression in *Drosophila melanogaster*. Genetics 177(2):801–808.

Yen JH, Barr AR. 1971. New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens*. Nature

Zabalou S, et al. 2004. *Wolbachia*-induced cytoplasmic incompatibility as a means for insect pest population control. Proc Natl Acad Sci U S A. 101(42):15042–15045.

Zug R, Hammerstein P, Cordaux R. 2012. Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. PLoS One 7(6):e38544.

Associate editor: Daniel Sloan