Title: Evolution-guided mutagenesis of the cytoplasmic incompatibility proteins: Identifying CifA's complex functional repertoire and new essential regions in CifB

Short Title: Mutagenesis of CI proteins

Authors: J. Dylan Shropshire[1,2*], Mahip Kalra[1,2], and Seth R. Bordenstein[1,2,3,4*]

[1]Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA

[2]Vanderbilt Microbiome Initiative, Vanderbilt University, Nashville, TN 37235, USA

[3]Department of Pathology, Microbiology, and Immunology, Vanderbilt University, Nashville, TN 37235, USA

[4]Vanderbilt Institute for Infection, Immunology, and Inflammation, Vanderbilt University Medical Center, Nashville, TN 37235, USA

*Correspondence to:
J. Dylan Shropshire, Nashville, TN, 37235, Phone 423.930.6292, e-mail shropxp@gmail.com
Seth R. Bordenstein, Nashville, TN, 37235, Phone 615.322.9087, e-mail s.bordenstein@vanderbilt.edu

ORCID iD: J. Dylan Shropshire (https://orcid.org/0000-0003-4221-2178), Seth R. Bordenstein (https://orcid.org/0000-0001-7346-0954)

Keywords: cytoplasmic incompatibility, *Wolbachia*, prophage WO, *Drosophila melanogaster*, Two-by-One model

Abstract

Wolbachia are the world's most common, maternally-inherited, arthropod endosymbionts. Their worldwide distribution is due in part to a selfish drive system termed cytoplasmic incompatibility (CI) that confers a relative fitness advantage to females that transmit *Wolbachia* to their offspring. CI results in embryonic death when infected males mate with uninfected females but not infected females. Under the Two-by-One genetic model of CI, males expressing the two phage WO proteins CifA and CifB cause CI, and females expressing CifA rescue CI. While each protein is predicted to harbor three functional domains, there is no knowledge on how sites across these Cif domains, rather than in any one particular domain, contribute to CI and rescue. Here, we use evolution-guided, substitution mutagenesis of conserved amino acids across the Cif proteins, coupled with transgenic expression in uninfected *Drosophila melanogaster*, to determine the functional impacts of conserved residues evolving mostly under purifying selection. We report that amino acids in CifA's N-terminal unannotated region and annotated catalase-related domain are important for both complete CI and rescue, whereas C-terminal residues in CifA's putative domain of unknown function are solely important for CI. Moreover, conserved CifB amino acids in the predicted nucleases, peptidase, and unannotated regions are essential for CI. Taken together, these findings indicate that (i) all CifA amino acids determined to be involved in rescue are correspondingly involved in CI, (ii) an additional set of CifA amino acids are uniquely important in CI, and (iii) CifB amino acids across the protein, rather than in one particular domain, are all essential for CI. We discuss how these findings advance an expanded view of Cif protein evolution and function, inform the mechanistic and biochemical bases of Cif-induced CI/rescue, and continue to substantiate the Two-by-One genetic model of CI.

Article summary

Wolbachia are maternally-transmitted, intracellular bacteria that occur in approximately half of arthropod species worldwide. They can spread rapidly though host populations via the cytoplasmic incompatibility (CI) drive system. CI causes embryonic death when infected males mate with infected females, but offspring of infected females are rescued. Two proteins, CifA and CifB, underlie the genetic basis of CI and rescue, but how amino acid sites across these proteins contribute to CI and/or rescue remain unknown. Here, we employed evolution-guided, combinatorial mutagenesis on conserved amino acids to understand their relative contributions to CI and rescue. The results of this study reveal a phenotypic complexity underlying the expression of these proteins and provide relevance to the biochemical and mechanistic bases of CI and rescue.

Introduction

Wolbachia are maternally-inherited, intracellular α-Proteobacteria that occur in 40-65% of all arthropod species (Charlesworth et al., 2019; Hilgenboecker et al., 2008; Weinert et al., 2015; Zug and Hammerstein, 2012). Residing in the cells of reproductive tissues, *Wolbachia* commonly cause a selfish drive system, cytoplasmic incompatibility (CI), that yields a relative advantage to females that transmit *Wolbachia* and thus increases *Wolbachia*'s rate of spread through the matriline (Rasgon, 2008; Turelli, 1994). CI causes embryonic death when *Wolbachia*-infected males mate with uninfected females and is rescued when the female is infected with a compatible *Wolbachia* strain (Fig. 1A) (LePage and Bordenstein, 2013; Serbus et al., 2008; Taylor et al., 2018). CI can act as a form of reproductive isolation between populations of different infection states (Bordenstein et al., 2001; Brucker and Bordenstein, 2012; Jaenike et al., 2006; Shropshire and Bordenstein, 2016). Additionally, this drive system has brought *Wolbachia* to the forefront of vector control efforts to combat Zika and dengue viruses because *w*Mel *Wolbachia* from *Drosophila melanogaster* flies confer resistance to RNA arboviruses when transinfected into *Aedes* mosquitoes (Aliota et al., 2016; Caragata et al., 2016; Kittayapong et al., 2018; Moreira et al., 2009; O'Connor et al., 2012; O'Neill, 2018; X. Zheng et al., 2019). Notably, *w*Mel-induced CI is the focus of this study.

CI's microbial genetic basis can be described by the Two-by-One genetic model (Fig. 1B) (Shropshire and Bordenstein, 2019) where the two phage WO genes *cifA* and *cifB* cause CI when dually expressed in testes (Beckmann et al., 2017; LePage et al., 2017; Shropshire and Bordenstein, 2019), and *cifA* rescues CI when singly expressed in ovaries (Chen et al., 2019; Shropshire et al., 2018; Shropshire and Bordenstein, 2019). The Cif proteins are divided into at least five phylogenetic clades, referred to as Types 1 – 5 (Bing et al., 2020; LePage et al., 2017; Lindsey et al., 2018). To date, this genetic model has been well supported using Type 1 *cif* genes from *w*Mel (LePage et al., 2017; Shropshire et al., 2018; Shropshire and Bordenstein, 2019) and is consistent with results with Type 1 and 4 genes from *w*Pip of *Culex pipiens* mosquitos (Beckmann et al., 2019b; Chen et al., 2019). However, the mechanism underlying CifA;B-induced CI and CifA-induced rescue remains mostly unresolved and limited to *in vitro* assays and structural homology-based predictive annotations discussed below.

$CifA_{wMel}$ is weakly predicted to encode a catalase-related domain (catalase-rel), a domain of unknown function 3243 (DUF), and a sterile-like transcription factor domain (STE) (Fig. 1C) (Lindsey et al., 2018). Catalase-rel domains are predicted to catalyze the degradation of reactive oxygen species (ROS) (Guy et al., 2005; Loew, 1900). DUF has a distant homology to globin-like domains and Puf-family RNA-binding domains, which influence the stability of eukaryotic RNAs

(Kumar and Subramaniam, 2018; Nishanth and Simon, 2020). Finally, STE domains mediate transcriptional induction in yeast (Wong Sak Hoi and Dumas, 2010). Structural homology predictions identify the STE and Puf-family RNA-binding domains across four and five phylogenetic Cif Types respectively, whereas the catalase-rel domain is only annotated in Type 1 (Bing et al., 2020; Lindsey et al., 2018).

Conversely, CifB$_{w\text{Mel}}$ harbors two putative PD-(D/E)XK-like nuclease domains (PDDEXK), and a ubiquitin-like-specific protease 1 domain (Ulp1) (Fig. 1C) (Beckmann et al., 2017; LePage et al., 2017; Lindsey et al., 2018). The Ulp1 domain is restricted to Type 1 CifB and cleaves poly-ubiquitin chains *in vitro* (Beckmann et al., 2017). Since mutation of the CifB Ulp1's catalytic motif ablates CifB's ability to contribute to CI when transgenically expressed in *D. melanogaster*, the Ulp1 domain has previously been described as the "enzymatic warhead" for CI induction (Beckmann et al., 2017). The function of the remaining sequence, including the PDDEXK dimer of Type 1 CifB, has not been assessed. We previously demonstrated that this remaining sequence comprises a CifB central core region found throughout Cif phylogenetic types (LePage et al., 2017). Moreover, the PDDEXK domains are also annotated in all phylogenetic Cif clades (Bing et al., 2020; LePage et al., 2017; Lindsey et al., 2018), and *in vitro* assays indicate that the PDDEXK domains of Type 4 CifB$_{w\text{Pip}}$ nick DNA (Chen et al., 2019). As such, CifB's central core including the PDDEXK domains is likely important, if not one of the central aspects of its ability to contribute to CI. Finally, *in vitro* assays suggest that CifA and CifB can bind (Beckmann et al., 2017), but the importance of this binding *in vivo* and for phenotypic output remains unknown.

The aforementioned protein annotations and biochemical data represent a foundation to develop a more complete and nuanced understanding of how Cif proteins cause CI and rescue, but it remains unclear what kind of impact each domain and unannotated region have on the phenotypic output of these proteins. Here, we test the importance of conserved amino acids across the Cif proteins to CI and rescue via site-directed substitution mutagenesis and transgenic expression in *D. melanogaster*. We report three key findings. First, conserved sites in CifA's N-terminal unannotated region and the catalase-rel domain are important in both CI and rescue. Second, conserved sites in CifA's DUF are only involved in CI. Finally, all tested conserved sites in CifB are required for CI. Taken together, we identify sites in seven Cif mutants (both CifA and CifB) essential for complete CI, and determine that CifA's N-terminus is involved in both CI and rescue while the middle of the protein is only involved in CI. These results inform the mechanistic and biochemical basis of CI and rescue and lend further support for a Two-by-One genetic model where both CifA and CifB are functionally crucial for expression of CI.

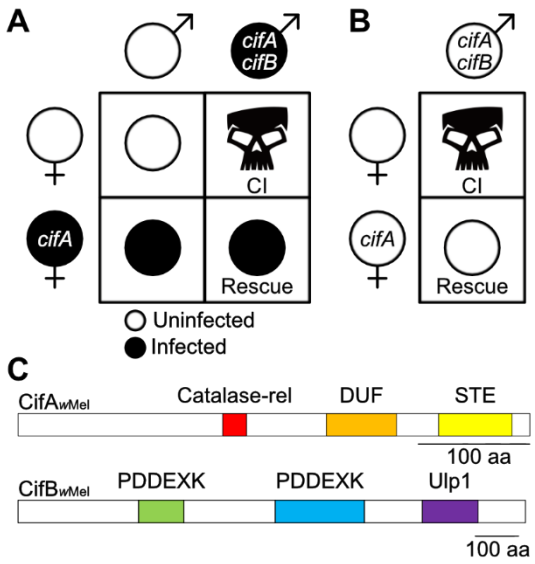

Figure 1. Cytoplasmic incompatibility, the Two-by-One genetic model, and Cif protein architecture. (A) CI is caused when *Wolbachia*-infected males expressing *cifA* and *cifB* mate with uninfected females. If females are infected and express cifA then offspring are rescued. (B) The Two-by-One genetic model indicates that males expressing cifA and cifB, even in the absence of an infection, can cause CI that can be rescued by uninfected females expressing cifA. (C) Schematic showing the protein architecture of CifA and CifB from the *w*Mel *Wolbachia* of *D. melanogaster*. Annotations are based on a prior structural homology-based analyses (Lindsey et al., 2018).

Results

CifA mutants impact CI and rescue

Since CifA does not have any putative catalytic motifs, we used a previous sequence analysis of conserved amino acid residues in an alignment of phylogenetically diverse CifA proteins (Lindsey et al., 2018) to select highly conserved sites across the protein for mutagenesis. $CifA_1$, $CifA_2$, $CifA_3$, and $CifA_4$ have combinatorial, alanine substitutions in the N-terminal unannotated region and putative catalase-rel, DUF, and STE domains, respectively (Fig. 2A). Alanine mutagenesis is used to analyze the importance of specific amino acids in protein sequences without contributing significant structural variation to the protein (Cunningham and Wells, 1989). We tested mutant *cifA* transgenes for their ability to (i) induce CI when dually expressed with wild type *cifB* in tests of uninfected males and (ii) rescue when singly expressed in ovaries of uninfected females. Using the GAL4-UAS system in *D. melanogaster* (Duffy, 2002), we predict that CI and/or rescue will not be recapitulated when CifA mutants are transgenically expressed if the sites mutated are crucial to the respective function. Notably, we use the *nos*-GAL4:VP16 driver throughout this study since it was previously shown to enable strong transgenic CI when dually expressing *cifA* and *cifB* transgenes (Shropshire and Bordenstein, 2019). Since

CI manifests as embryonic death, we measured the strength of CI induced under mutant transgenic expression by measuring the percentage of *D. melanogaster* embryos that hatch into larvae.

Consistent with prior studies (Shropshire and Bordenstein, 2019), dual *cifA;B* expression in males yielded nearly complete CI-defining embryonic death when mated to uninfected females (Mdn = 0% hatching) that was statistically comparable to *w*Mel-induced CI ($p > 0.99$), and it was rescued by *w*Mel-infected females (Mdn = 95.4% hatching) (Fig. 2B). Transgenic dual expression of either *cifA*$_1$*;B* (Mdn = 93.7% hatching; $p > 0.99$) or *cifA*$_3$*;B* (Mdn = 94.1% hatching; $p > 0.99$) in males crossed to uninfected females revealed no statistically significant difference in hatching relative to the compatible, rescue cross; thus, mutating conserved sites in CifA's unannotated region and putative DUF ablates CI. Conversely, transgenic expression of *cifA*$_4$*;cifB* caused hatch rates statistically comparable to *cifA;B*-induced CI (Mdn = 0% hatching; $p > 0.99$), suggesting that mutation of conserved sites in the putative STE did not impact *cifA*'s ability to contribute to CI. Finally, transgenic expression of *cifA*$_2$*;B* (Mdn=66.0%) yielded an intermediate phenotype whereby it was statistically different from both *cifA;B*-induced CI ($p = 0.0006$) and rescue of transgenic CI ($p = 0.0001$), indicating that the putative catalase-rel mutant induces a partial CI phenotype. Together, these results suggest the mutated sites in the unannotated region, catalase-rel, and DUF of CifA are important for CI-induction (Fig. 2B).

Next, we reciprocally tested if uninfected transgenic females singly expressing the same *cifA* mutants rescue *cifA;B*-induced CI (Fig. 2C). As above, dual *cifA;B* expressing males induced near-complete embryonic death consistent with strong CI (Mdn = 0% hatching), and this lethality could be rescued when the female expressed *cifA* (Mdn = 85.9% hatching). Transgenic expression of *cifA*$_1$ (Mdn = 0.00%; $p < 0.0001$) and *cifA*$_2$ (Mdn=27.6%; $p = 0.0390$), which failed to contribute to CI, also failed to rescue *cifA;B*-induced CI as compared to the standard transgenic rescue cross. Conversely, transgenic expression of *cifA*$_3$ (Mdn=91.2%; $P > 0.9999$) and *cifA*$_4$ (Mdn=97.6%; $P = 0.3039$) rescued *cifA;B*-induced CI at levels comparable to the standard transgenic rescue cross. These results suggest that the sites mutated in the unannotated and catalase-rel regions of CifA are important for rescue (Fig. 2C).

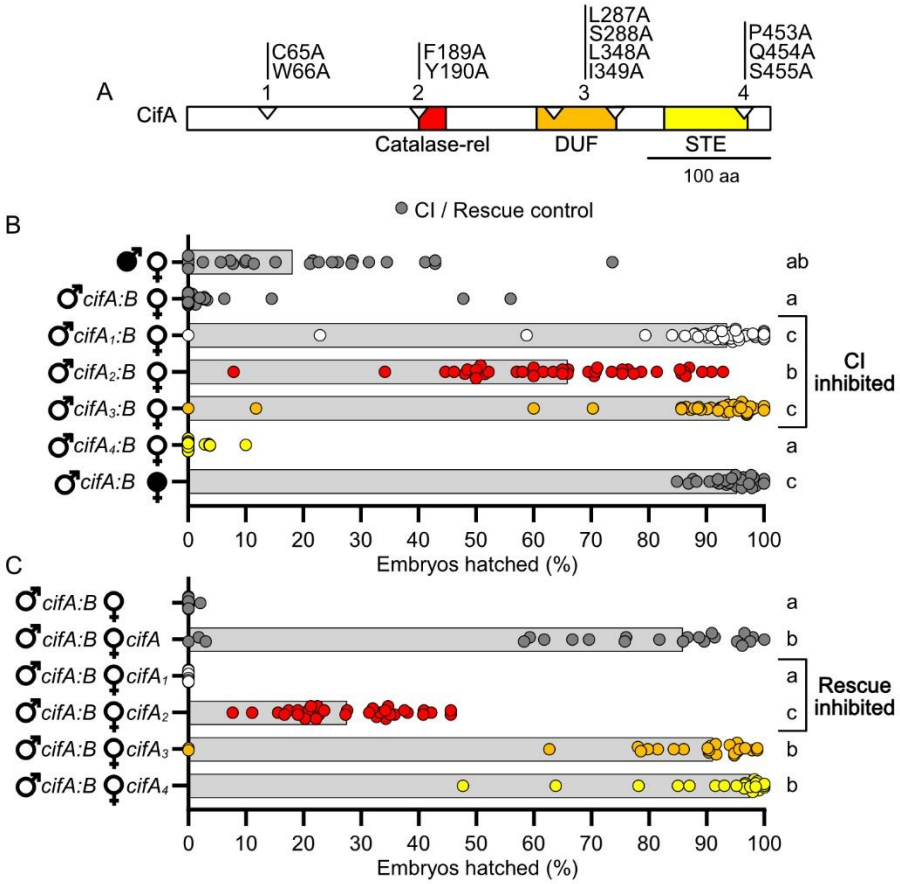

Figure 2. *cifA₁ and cifA₂* fail to cause or rescue CI, and *cifA₃* can rescue but fails to cause CI. (A) schematic showing the location of amino acid mutations in CifA relative to previously predicted domains (Lindsey et al., 2018). (B) Hatch rate experiment testing if *cifA* mutants can induce CI when dual expressed with *cifB* in uninfected males. (C) Hatch rate experiment testing if expressing *cifA* mutants can rescue transgenic CI when expressed in uninfected females. (B/C) Each dot represents the percent of embryos that hatched from a single male and female pair. Expressed genes are noted to the right of the corresponding sex. Gray bars represent median hatch rates for each cross and letters to the right indicate significant differences based on α = 0.05 calculated by Kruskal-Wallis and Dunn's test for multiple comparisons between all groups. Panel B was conducted three times and C was conducted twice. P-values are reported in Table S1.

CifB mutants ablate CI

Four CifB mutants were constructed based on a comparative sequence analysis of conserved residues (Lindsey et al., 2018). All CifB mutations are similarly alanine substitutions, with the exception of one glycine mutation of a conserved alanine (Fig. 3A). Glycine was chosen to replace alanine since it is comparably sized and would be less likely to impact protein structure than other amino acids. $CifB_1$, $CifB_2$, $CifB_3$, and $CifB_4$ have mutations in the N-terminal unannotated region, first PDDEXK, second PDDEXK, and Ulp1 respectively (Fig. 3A). The Ulp1 mutation is the same used previously to test for the catalytic activity of the Ulp1 domain (Beckmann et al., 2017). We predict that CI will not be recapitulated when CifB mutants are transgenically expressed if the sites mutated are crucially important for CI-induction. As with CifA

mutants above, we tested mutant CifB for their ability to induce CI when dually expressed with *cifA* in uninfected males.

As expected, dual *cifA;B* expression in uninfected males caused hatch rates statistically comparable to *w*Mel-induced CI ($p > 0.99$), and it could be rescued by *w*Mel-infected females (Mdn = 93.9% hatching). However, transgenic expression of *cifA;B*$_1$ (Mdn = 96.3%; $p < 0.0001$), *cifA;B*$_2$ (Mdn = 95.6%; $p < 0.0001$), *cifA;B*$_3$ (Mdn = 94.3%; $p < 0.0001$), and *cifA;B*$_4$ (Mdn = 93.0%; $p < 0.0001$) all failed to reduce hatch rates statistically comparable to *cifA;B*-induced CI (Mdn = 0.%). These results specify that all mutated conserved sites, rather than any one site or domain such as the previously reported catalytic site of Ulp1 (Beckmann et al., 2017), are important for CifB in CI-induction.

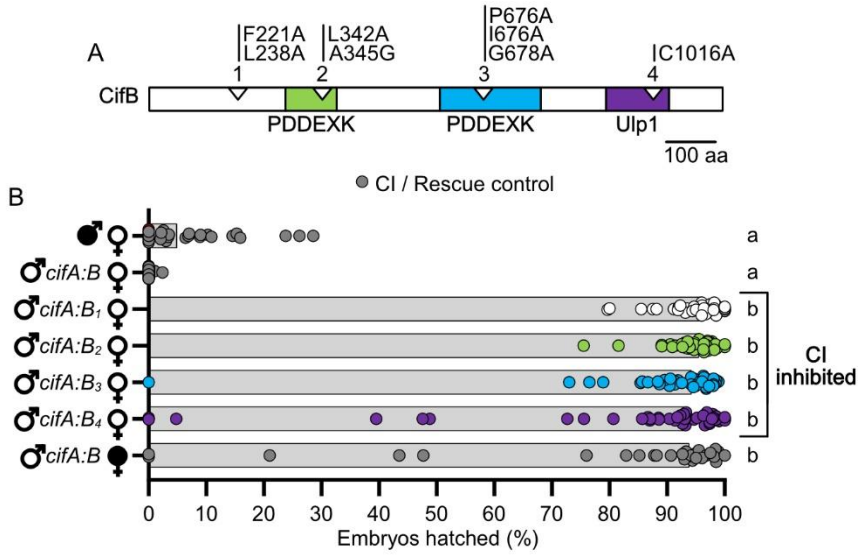

Figure 3. All *cifB* mutants fail to contribute to CI. (A) schematic showing the location of mutations in CifB relative to previously predicted domains (LePage et al., 2017; Lindsey et al., 2018). (B) Hatch rate experiment testing if *cifB* mutants can induce CI when dual expressed with *cifB* in uninfected males. Each dot represents the percent of embryos that hatched from a single male and female pair. Expressed genes are noted to the right of the corresponding sex. Gray bars represent median hatch rates for each cross and letters to the right indicate significant differences based on $\alpha = 0.05$ calculated by Kruskal-Wallis and Dunn's test for multiple comparisons between all groups. Panel B was conducted three times. P-values are reported in Table S1.

CifB selection analysis.

Prior selection analyses using a sliding window analysis indicated that CifA is under strong purifying selection, and CifA's N-terminal unannotated region and catalase-rel domain are under stronger purifying selection than C-terminal regions (Shropshire et al., 2018). Here, in order to assess how engineered mutations and phenotypes align with selective pressures, we replotted this analysis for CifA and, for the first time, applied a sliding window analysis of K_a and K_s (SWAKK) to assess the selective pressure on CifB (Fig. 4). SWAKK calculates the ratio of non-

synonymous to synonymous substitution rates (K_a/K_s) in a pairwise alignment. Codons with ratios below one are considered under purifying selection, and those above one are under positive selection. In both cases, pairwise nucleotide alignments were *cif*$_{wMel}$ relative to *cif*$_{wHa}$, which are modestly divergent Type 1 *cif* variants. For both CifA and CifB, we then plotted the amino acid substitutions to assess if they occur in regions of purifying or positive selection.

For CifA, all mutant sites were in areas of purifying selection (Fig. 4). However, the strongest purifying selection was surrounding the codons we mutated in the N-terminal unannotated region and catalase-rel domain. Intriguingly, the sites mutated near CifA's STE are under strong purifying selection, but the rest of the domain evolves neutrally. These results are in-line with our observations that the mutations in the N-terminal unannotated region and catalase-rel domain are involved in both CI and rescue, whereas weaker selection may act on the DUF domain which based on the results here are only involved in CI. Additionally, as with CifA, the majority of CifB is under strong purifying selection (Fig. 4). However, there are two clusters of amino acids that appear to be under strong positive selection. These residues are on the C-terminal side of CifB's second PDDEXK domain and in the unannotated region between the PDDEXK domain and Ulp1 domain. $CifB_1$'s mutations fall within regions under neutral pressures, whereas the other CifB mutants are in regions under strong purifying selection.

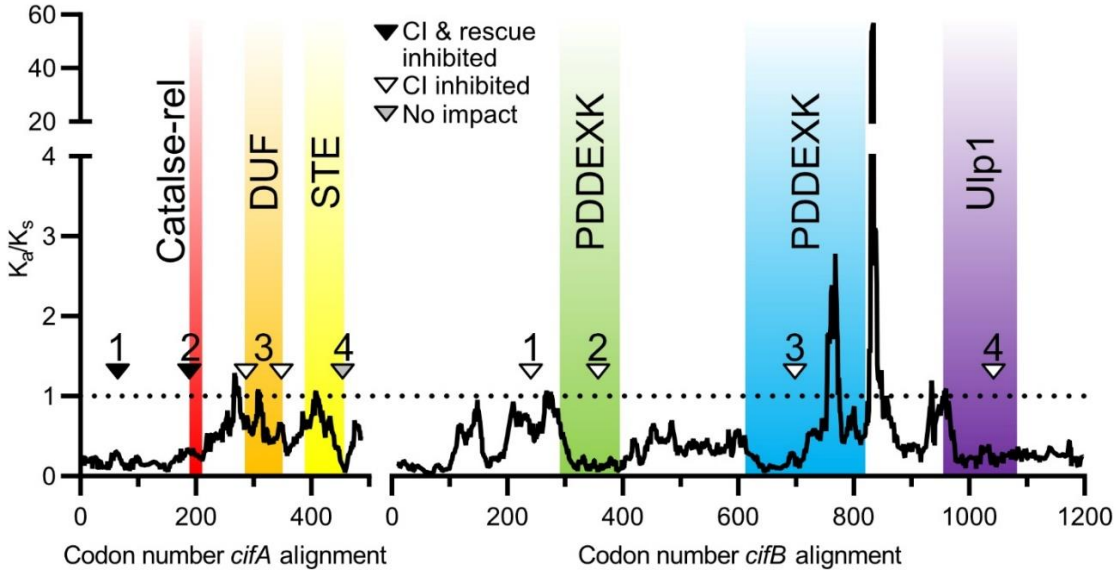

Figure 4. Sliding window analyses reveal selective pressures on regions surrounding evolution-guided mutagenesis. Analyses are based on pairwise alignments of *cifA* and *cifB* variants of *w*Mel and *w*Ha where the ratio of non-synonymous to synonymous substitutions (K_a/K_s) are calculated along a 25 amino acid sliding window. The analysis of CifA was previously reported (Shropshire et al., 2018). Both proteins are largely under strong purifying selection, but CifB has two regions of positive selection. Sites mutated in $CifA_1$, $CifA_2$, $CifA_4$, $CifB_2$, $CifB_3$, and $CifB_4$ are in regions of purifying selection. Sites mutated in $CifA_3$ and $CifB_1$ appear to be under neutral pressures or slight purifying selection. Triangles on protein schematics represent a selection of mutated amino acids. The specific amino acids mutated are further illustrated in Fig. 2A and 3A for CifA and CifB respectively. Annotations are from prior studies and based on structural-homology based analyses (Lindsey et al., 2018).

Cif structural predictions

There are numerous ways to interpret the impact of a mutation on a protein's function. These can include non-exclusive changes to catalytic motifs, ligand binding sites, or changes in local or global structures that ablate, enhance, or otherwise modify the phenotypic output of the protein. The $CifB_4$ mutation in the Ulp1 domain involves a putative cysteine catalytic motif common to deubiquitinase domains (Beckmann et al., 2017). Otherwise, no other catalytic motifs or binding sites have been identified in $CifA_{wMel}$ or $CifB_{wMel}$. As such, we aimed to investigate the impact of these mutations on the structure of CifA and CifB proteins.

The Iterative Threading ASSEmbly Refinement (I-TASSER) webserver was used to generate a list of structural homologs from the protein databank (PDB) for each wild-type and Cif mutant protein and construct structural models based on these hits (Zhang, 2009). The shared and unique PDB hits for wild-type and mutant proteins are summarized in Fig. S1A and detailed in Table S2. The top 10 PDB hits for each protein were used to create structural models (Fig. S1B). Each model is generated with confidence measures in the form of C-scores and TM-scores. C-scores range from -5 to 2 where 2 is the highest confidence, and TM-scores range from 0-1 where 1 is the highest confidence (Zhang, 2009). The similarity between wild-type and mutant structures was then assessed using the Alignment plugin in PyMOL 2.3.2 which provides values for the root-mean-square deviation (RMSD) of atomic positions. Higher RMSDs indicates a greater distance between the atoms of mutant proteins superimposed on the wild-type protein. Structural models were generated for CifA (C-score = -2.74;TM = 0.4±0.13), $CifA_1$ (C-score = -1.42; TM = 0.54±0.15; RMSD = 19.7), $CifA_2$ (C-score = -2.84;TM = 0.39±0.13; RMSD = 1.9), $CifA_3$ (C-score = -1.39;TM = 0.54±0.15; RMSD = 20.2), and $CifA_4$ (C-score = -1.43;TM = 0.54±0.15; RMSD = 19.8) (Fig. S1B). These low C-scores and TM-scores indicate that the I-TASSER predictions for CifA are not robust, and that variation in structure between wild-type and mutant proteins could be the result of poor threading templates. However, these results suggest that CifA is structurally most comparable to $CifA_2$ while the other structures are predicted to change to comparable degrees. Crucially, since the most divergent model, $CifA_3$, remains rescue-capable and the second most divergent model $CifA_4$ remains functional in both CI and rescue it is unlikely that predictions of global structural variation assists an understanding of phenotypic ablation in CifA.

I-TASSER was also used to identify PDB hits and create structures for wild-type and mutant CifB. The shared and unique PDB hits for wild-type and mutant proteins are summarized in Fig. S1C and detailed in Table S2. As above, I-TASSER protein structures were then created based on these threading templates and compared for RMSD. Structural models (Fig. S1D) were

generated for CifB (C-score = -1.02, TM-score = 0.59±14), $CifB_1$ (C-score = -0.68, TM-score = 0.63±14; RMSD = 1.7), $CifB_2$ (C-score = -1.03, TM-score = 0.58±14; RMSD = 10.6), $CifB_3$ (C-score = -1.07, TM-score = 0.58±14; RMSD = 2.2), $CifB_4$ (C-score = -0.65, TM-score = 0.63±14; RMSD = 2.2) (Fig. S1D). Together, these results suggest that all CifB mutants are structurally comparable to the wild-type protein. As with CifA, it remains unknown how small effects in protein structure may influence phenotypic ablation.

Discussion

CI's genetic basis involves a Two-by-One genetic model whereby *cifA;B* expressing males cause CI and *cifA* expressing females rescue CI (Beckmann et al., 2017, 2019b; LePage et al., 2017; Shropshire et al., 2018; Shropshire and Bordenstein, 2019). However, the mechanistic basis of *cifA;B*-induced CI and *cifA*-induced rescue remains largely unresolved. Here, we test the phenotypic impact of conserved sites (Lindsey et al., 2018) to CI and rescue across the Cif_{wMel} proteins *in vivo* using site-directed mutagenesis and transgenic expression in *D. melanogaster*. We discuss the relevance of these findings, particularly the complex functional repertoire of CifA, to the Toxin-Antidote (TA) and Host Modification (HM) models of CI (Beckmann et al., 2019a; Shropshire et al., 2019) and the molecular basis of CifA;B-induced CI and CifA-induced rescue.

Crucially, phenotypic ablation caused by site-directed substitution mutagenesis has multiple possible and non-exclusive interpretations. First, a mutated site may be part of a catalytic motif. If so, mutagenesis may result in a significant inhibition of enzymatic activity related to CI and/or rescue induction. Second, the mutated site may be crucial for binding to host ligands, binding between CifA and CifB, or binding to nucleotide products. Additionally, amino acid substitutions may impact protein structure which in turn could also ablate function. Global folding abnormalities may block the enzymatic function of trans-acting domains or prevent binding elsewhere in the protein. Local or non-local changes to the structure can inhibit processes close to the mutated site in tertiary structure (Mi et al., 2019; Yang et al., 2020). These hypotheses are applicable to all mutant variants. Below, we place our mutagenesis work in the context of prior *in vivo* assays and bioinformatics to develop hypotheses regarding the mechanistic basis of CI and rescue.

First and foremost, our results indicate that CifA has several amino acids with overlapping functions in CI and rescue, as well as amino acids of specific importance to CI. Thus, it is plausible that, CifA's functional role in CI is comparable to that in rescue. Among the two sets of mechanistic models for CI, HM-based models predict that expression of CifA and CifB cause CI through host modifications during spermatogenesis, and rescue occurs by reversing those modifications in the

329 embryo (Shropshire et al., 2019). In the context of these results, CifA's role in modifying host
330 factors in the testes may be comparable to its activity in rescue. Indeed, the HM-based mistiming
331 model posits that CI causes a delay in nuclear apposition during the first mitosis due to slowed
332 development of the male pronucleus, and rescue occurs when the development of the female
333 pronucleus is comparably delayed (Ferree and Sullivan, 2006; Tram and Sullivan, 2002). Thus,
334 under the mistiming model, CifA's overlapping function in CI and rescue could indicate that it is
335 the primary driver of CI-induction and rescue, whereas CifB and CI-specific sites in CifA may
336 provide adjunct functions necessary to access those targets when expressed in the testes.
337 Conversely, the TA model predicts that CifB is paternally-transferred and the primary factor that
338 causes an embryonic toxicity, and CifA is necessary to prevent self-induced toxicity in the testes
339 and to rescue CifB-toxicity in the embryo (Beckmann et al., 2019a). Unlike the HM model, the TA
340 model predicts that CifA must function exactly as an antidote when expressed in testes and
341 embryos. Thus, our data are also in-line with these expectations, but as with the HM model, the
342 TA model must be expanded to explain why sites in CifA's DUF domain are only essential for CI-
343 induction if CifA's primary function is as an antidote in both tissues. Crucially, there is no evidence
344 in the literature that CifB is transferred to the embryo, and we have proposed a HM model wherein
345 CifA is both a primary inducer of CI in testes and rescue in embryos, while CifB is an accessory
346 protein; this scenario enables a simple, one-step mutation scenario to the evolution of bidirectional
347 CI between different strains of *Wolbachia* (Shropshire et al., 2018). More work will be necessary
348 to understand CifA's cell biology and biochemistry to support it.

349 In the context of CifA, mutations in the N-terminal region completely ablate both CI and
350 rescue. Intriguingly, prior selection analyses revealed that this region is under stronger purifying
351 selection than the C-terminal end of the protein, suggesting that conservation in CifA's N-terminus
352 may have a strong impact on its phenotypic output (Shropshire et al., 2018). Given that these N-
353 terminal regions in both CifA and CifB are unannotated, we cannot yet provide a specific
354 explanation for why these regions ablate CI and rescue. However, it is notable that N-terminal
355 regions in *w*Pip homologs of both CifA and CifB are predicted to encode ankyrin-interacting
356 domains (Bonneau et al., 2018). Thus, these mutations may be responsible for ablating the
357 binding capacity of Cif proteins to host ligands through either sequence or structural modifications
358 of the protein, which requires further study.

359 When conserved sites within CifA's putative catalase-rel domain were mutated and dually
360 expressed with CifB in males, 34.05% of embryos died indicating that CI capability was inhibited
361 but not ablated. Similarly, when expressed in females crossed to CI-inducing CifA;B males, only
362 34.05% of embryos survived, indicating that rescue was also significantly weaker. Thus, these

amino acids are important for causing complete CI and rescue phenotypes. Catalases are enzymes that are involved in the decomposition of hydrogen peroxide and protect cells from reactive oxygen species (ROS) damage (Loew, 1900). Some catalase-like domains are involved in host immune pathways that use ROS to combat disease (Govind, 2008; Zug and Hammerstein, 2015), and high levels of ROS can cause male infertility in organisms as diverse as *Drosophila* and humans (Homa et al., 2015; Yu and Huang, 2015). Notably, while CifA is annotated with a catalase-rel domain (Lindsey et al., 2018), the closest sequence homolog is from *Helicobacter pylori*, which shares only ~22% sequence identity (Guy et al., 2005) and has no obvious active sites (Lindsey et al., 2018). Thus, it is unknown if CifA's catalase-rel is capable of degrading ROS, but it may otherwise attract or interact with ROS while not degrading them. For example, oxidative posttranslational modifications (PTM) can shift phenotypic output (Cai and Yan, 2013). CifA's catalase-rel may attract ROS, enabling PTM of itself, CifB, or other host targets. Since oxidative PTMs can be reversible (Cai and Yan, 2013), rescue may in part occur through the removal of these PTMs in the embryo. Alternatively, CifA may help to localize ROS to host targets to induce oxidative damage or otherwise modify host targets.

Moreover, *Wolbachia* presence is correlated with increases in ROS in *D. melanogaster, D. simulans*, *A. albopictus*, *A. polynesiensis*, and *T. urticae* (Brennan et al., 2012, 2008; Zug and Hammerstein, 2015), yielding additional support for CifA's function as a catalase-related protein (Lindsey et al., 2018). *Wolbachia*-induced increases in ROS levels correlate with DNA damage in *D. simulans* spermatocytes (Brennan et al., 2012) and an increase in lipid hydroperoxides in *D. melanogaster,* which are markers for ROS-induced oxidative damage (Driver et al., 2004). Intriguingly, the immune-related gene *kenny* (*key*) is upregulated in *Wolbachia*-infected *D. melanogaster*, and experimental upregulation of *key* in uninfected male flies yielded increased ROS levels, DNA damage, and decreased hatching that can be rescued when mated to infected females (Biwot et al., 2019). Together, these data support a role for ROS in CI's mechanism, but more work is necessary to link CifA's catalase-rel domain to ROS variation and determine if ROS are directly responsible for CI/rescue-induction or are otherwise a symptom of other modifications in gametogenesis.

The DUF in CifA shares distant homology to Puf-family RNA-binding proteins and is the only putative domain shared in all five CifA clades (Bing et al., 2020). Notably, mutating conserved residues in CifA's DUF domain revealed their specific importance to CI-induction but not for rescue. As such, this was the only domain in CifA wherein mutated sites were differentially important between the two phenotypes. RNA-binding proteins are important in transcriptional regulation and can influence the stability, localization, and translation of bound RNA. Puf-family

RNA-binding proteins typically influence the stability of mRNAs involved in cell maintenance, embryonic development, and other processes (Forbes and Lehmann, 1998; Macdonald, 1992; Parisi and Lin, 1999). For example, the *Drosophila* Puf-family RNA *Pumilio* (*pum*) is crucially involved in the establishment of patterning and abdominal segmentation in early embryonic development by suppressing the translation of maternal hunchback RNA in the *Drosophila* embryo (Forbes and Lehmann, 1998; Weidmann and Goldstrohm, 2012). Moreover, *pum* in spermatogenesis negatively regulates the expression of p53 which is involved in DNA repair and apoptosis, increases apoptosis, and reduces sperm production and fertility (Chen et al., 2012). Intriguingly, mitochondrial protein p32 is a candidate suppressor of CI based on *in vitro* pull-down assays using *Drosophila* lysates, and p32 regulates p53 activation (Ghate et al., 2019), suggesting that *pum* in spermatogenesis may influence similar pathways in CI. Additionally, *Wolbachia* infection has been shown to have a considerable impact on the fly transcriptome including sRNA profiles (Baião et al., 2019; Pinto et al., 2013; Y. Zheng et al., 2019). On their own, these correlations do not sufficiently link transcription with CI. However, as described above, *key* is significantly upregulated and causes rescuable hatch rate defects when experimentally overexpressed (Biwot et al., 2019). Additionally, *Wolbachia* upregulate the sRNA nov-miR-12 which negatively regulates *pipsqueak* (*psq*), a DNA-binding protein that impacts chromatin structure (Horowitz and Berg, 1996; Siegmund and Lehmann, 2002), and knockdown of *psq* causes CI-like embryonic abnormalities and hatch rates in *D. melanogaster* (Y. Zheng et al., 2019). Thus, CifA's DUF may influence the expression of RNAs involved in the CI pathway, and by mutating the conserved site it may ablate the domain's ability to regulate these RNAs. It is also important to note that since the DUF mutant only prevented CifA from contributing to CI, it supports prior hypotheses that CifA has distinct mechanistic input to CI and rescue (Shropshire and Bordenstein, 2019). The phenotypic plasticity of CifA may be caused by distinct protein conformations in testes and ovaries or DUF-associated targets may only be present in testes. More work will be necessary to confirm that CifA can bind RNAs and what impact this binding has on downstream processes.

The final CifA domain shares homology to STE transcription factor proteins which are found predominantly in fungi and encode a sequence-specific DNA-binding motif that influences yeast reproduction through pheromone-responsive elements (Wong Sak Hoi and Dumas, 2010). Mutation of conserved sites within the STE had no impact on either CI or rescue. This was surprising since the STE domain appeared structurally conserved across four of the five CifA phylogenetic Types (Bing et al., 2020; Lindsey et al., 2018). However, while the sites mutated in the domain were conserved, the remainder of the domain appears to evolving mostly neutrally

(Shropshire et al., 2018), suggesting the domain may be of lesser importance than N-terminal regions. Alternatively, since transgenes are expressed via host transcription and translation machinery, transgenic expression bypasses the need to export the proteins outside of *Wolbachia* and through the host-derived membranes that surround *Wolbachia* within the cell (Cho et al., 2011; Fattouh et al., 2019). As such, it is possible that the STE domain has an essential functional role in translation initiation and/or in protein export within *Wolbachia*, which may or may not be related to CI and/or rescue induction. Additional research will be necessary to determine if any component of the STE domain is necessary for CI and whether this domain is essential when expressed inside *Wolbachia*.

Type I CifB have two domains with homologs in the PDDEXK nuclease family, but they do not encode a canonical PD-(D/E)XK catalytic motif (Knizewski et al., 2007; Lindsey et al., 2018). This nuclease family is heavily involved in DNA restriction, repair, recombination, and binding (Knizewski et al., 2007). Mutations in conserved sites in either domain of Type I $CifB_{wMel}$ ablated CI phenotypes. Interestingly, homologs of CifB proteins from all five phylogenetic Types harbor putative nuclease domains (Lindsey et al., 2018), and the Type IV CifB have functional nucleases with canonical catalytic motifs and can induce CI upon dual expression with CifA (Chen et al., 2019). However, it remains biochemically unclear how these DNA nicks contribute to wild-type CI induction, and how this can be rescued by CifA expressing females. Moreover, while we show here that mutating conserved residues in either PDDEXK domain ablates CI induction, this does not confirm its role as a nuclease. For instance, these domains may be essential for the localization of Cif proteins to host DNA or other host targets. More work will be necessary to determine if CifB's PDDEXK domains are indeed active nucleases and why mutating these conserved sites ablates CI. Though, we note that ablation of CI by site-specific mutagenesis across the Cif proteins highlights the utility of a mechanistically-agnostic, gene nomenclature, like the CI factor (*cif*) gene nomenclature, as it is increasingly clear that CifB's role as a deubiquitinase or nuclease alone does not define its role in CI (Beckmann et al., 2019a; Shropshire et al., 2019).

CifB's final domain is a Ulp1 domain that contains the only known catalytic motif within the Cif proteins and is responsible for the deubiquitinase activity observed *in vitro* (Beckmann et al., 2017). Previous reports show that mutating the conserved cysteine active site ablates CI function in $CifB_{wPip}$ (Beckmann et al., 2017). Here, we confirm that mutating the same cysteine active site ablated CI in the $CifB_{wMel}$ protein. However, it is premature to claim that the Ulp1 domain is a "catalytic warhead" for CI (Beckmann et al., 2017) because several sites, when mutated, ablate the CI phenotype. Instead, it is evident that this site in the Ulp1 domain plays an important role in CI-induction but, when analyzed in context of results from mutating other regions, it is not the only

crucially important component of the protein. More work will be necessary to dissect the relative importance of the unannotated region, the nuclease domains, and the Ulp1 in CI's mechanism.

In conclusion, we report conserved amino acids in CifA and CifB that are essential for CI and rescue phenotypes. For CifA, conserved sites in the unannotated region and catalase-rel domain were important for CifA-induced CI and rescue, while the mutated sites in the DUF was specifically important to CI. For CifB, mutating conserved sites in an unannotated region, both PDDEXK nuclease domains and the Ulp1 domain were important in CifB-induced CI. These works provide additional support for the necessity of expressing both CifA and CifB proteins to cause CI and the importance of CifA's complex functional repertoire and new essential regions in CifB to the genotype-phenotype relationship and mechanism underpinning CI.

Materials and methods

Creating transgenic flies.

cifA$_1$ and *cifB$_1$* mutant transgene variants were synthesized *de novo* at GenScript and cloned into a pUC57 plasmid. Site-directed mutagenesis was then performed by GenScript to produce the remaining three mutant variants of each gene (Fig 2A, 3A). UAS transgenic *cifA* and *cifB* mutant flies were then generated following previously described protocols (LePage et al., 2017). Briefly, each gene was subcloned into the pTIGER plasmid, which is a pUASp-based vector designed for germline expression. *cifA* and *cifB* transgenes were integrated into the *attp40* and *attp2* attachment sites in the *D. melanogaster* genome using PhiC31 integrase via embryonic injections at BestGene (LePage et al., 2017). All wild-type and transgenic nucleotide and amino acid sequences are reported in Table S3.

Fly rearing and strains.

D. melanogaster stocks *y^{1}w** (BDSC 1495), *nos*-GAL4:VP16 (BDSC 4937), and UAS transgenic lines homozygous for *cifA*, *cifA* mutants, *cifB*, *cifB* mutants , *cifA*;*B*, and lines dual homozygous for *cifA* or *cifB* mutants with wild-type counterparts were maintained on a 12-hour light/dark cycle at 25^{o}C on 50mL of standard media. Dual transgenic lines were generated through standard genetic crossings and were all homozygous viable. Uninfected lines were produced by tetracycline treatment as previously described (LePage et al., 2017). Infection status for all lines was regularly confirmed by PCR using Wolb_F and Wolb_R3 primers (Casiraghi et al., 2005). Genotyping was confirmed by PCR and Sanger sequencing using the primers in Table S4.

CI measurement assays.

CI was measured using hatch rate assays. To control for the paternal grandmother age effect on CI (Layton et al., 2019), virgin *nos*-GAL4:VP16 females were collected for the first 3 days of emergence and aged 9-11 days before crossing to nonvirgin UAS transgenic males. Collections for maternal and paternal lineages were separated by a 7-day period. Individual male and female mating occurred in 8-oz *Drosophila* stock bottles with a grape-juice agar plate smeared with yeast and secured to the opening of each bottle with tape. Only the first emerging and youngest males were used to control for the younger brother effect and age effects on CI (Reynolds and Hoffmann, 2002; Yamada et al., 2007). Grape-juice agar plates were produced as previously described (LePage et al., 2017). The flies and bottles were incubated at 25^{o}C for 24 hours, at which time the grape plates were replaced with fresh plates and stored for an additional 24 hours. After this, the initial number of embryos on each plate were counted. The plates were incubated at 25^{o}C and after 30 hours, the number of unhatched embryos was counted. The percentage of embryos that hatched was calculated by dividing the number of hatched embryos by the total number of embryos and multiplying by 100. Plates with fewer than 25 embryos were excluded from analysis as previously described (LePage et al., 2017; Shropshire et al., 2018).

Selection analysis.

Selection analysis of CifB was conducted using a sliding window analysis of K_a and K_s (SWAKK). The SWAKK 2.1 webserver first generated a 1196 codon alignment (with gaps) between $cifB_{wMel}$ and $cifB_{wHa}$ using ClustalW2 and then calculated the Ka/Ks ratio across the gene alignment using a sliding window of 25 codons and a jump size of 1 codon for SWAKK. The *cifA* SWAKK analysis was previously published and is based on the same parameters described above for cifB (Shropshire et al., 2018).

Predicting mutational impact on protein structure.

The effect of mutations on protein structure was evaluated with the I-TASSER protein prediction tool (Zhang, 2009). I-TASSER generated protein tertiary structure predictions for Cif proteins and their mutants using the on-line server with default settings. Structures are build based on the top ten hits generated by querying the PDB. Hits were provided Z-scores that characterize the similarity to the query sequence. Higher Z-scores represent more confident matches. Z-scores are reported in the source data provided with this manuscript and PDB hits are detailed in Table S2. C-scores and TM-scores were generated for each tertiary structure. C-scores range from -5 to 2 where 2 is the highest confidence. TM-scores range from 0-1 where 1 is the highest confidence.

Statistical analysis.

All statistical analyses for hatch rates were conducted in GraphPad Prism 8. Hatch rate statistical comparisons were made using Kruskal-Wallis followed by a Dunn's multiple comparison test. All p-values from statistical comparisons are provided in Table S1. Figure aesthetics were edited using Affinity Designer.

Data availability

All data generated in this study are available in the supplement of this manuscript.

Acknowledgments

We thank members of the Bordenstein lab, especially Brittany Leigh, for helpful comments and critiques during the course of this study. This work was supported by National Institutes of Health awards R01 AI132581 and R01 AI143725, National Science Foundation award IOS 1456778, and the Vanderbilt Microbiome Initiative to S.R.B., and a National Science Foundation Graduate Research Fellowship DGE-144519 to J.D.S. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Institutes of Health, the National Science Foundation, or Vanderbilt University.

Competing interests

J.D.S. and S.R.B. are listed as inventors on a provisional patent relevant to this work. S.R.B. is a coinventor on two other pending patents related to controlling arthropods.

Author roles

J.D.S: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing. **M.K.:** Formal Analysis, Investigation, Writing – Review & Editing. **S.R.B.:** Conceptualization, Funding Acquisition, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

564

References

Aliota, M.T., Peinado, S.A., Velez, I.D., Osorio, J.E., 2016. The *w*Mel strain of *Wolbachia* reduces transmission of Zika virus by *Aedes aegypti*. Sci. Rep. 6, 28792. https://doi.org/10.1038/srep28792

Baião, G.C., Schneider, D.I., Miller, W.J., Klasson, L., 2019. The effect of *Wolbachia* on gene expression in *Drosophila paulistorum* and its implications for symbiont-induced host speciation. BMC Genomics 20, 465. https://doi.org/10.1186/s12864-019-5816-9

Beckmann, J., Bonneau, M., Chen, H., Hochstrasser, M., Poinsot, D., Merçot, H., Weill, M., Sicard, M., Charlat, S., 2019a. The Toxin–Antidote Model of Cytoplasmic Incompatibility: Genetics and Evolutionary Implications. Trends Genet. https://doi.org/10.1016/j.tig.2018.12.004

Beckmann, J., Ronau, J.A., Hochstrasser, M., 2017. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. Nat. Microbiol. 2, 17007. https://doi.org/10.1038/nmicrobiol.2017.7

Beckmann, J., Sharma, G.D., Mendez, L., Chen, H., Hochstrasser, M., 2019b. The *Wolbachia* cytoplasmic incompatibility enzyme CidB targets nuclear import and protamine-histone exchange factors. eLife 8, e50026. https://doi.org/10.7554/eLife.50026

Bing, X.-L., Zhao, D.-S., Sun, J.-T., Zhang, K.-J., Hong, X.-Y., 2020. Genomic analysis of *Wolbachia* from *Laodelphax striatellus* (Delphacidae, Hemiptera) reveals insights into its "Jekyll and Hyde" mode of infection pattern. Genome Biol. Evol. https://doi.org/10.1093/gbe/evaa006

Biwot, J.C., Zhang, H.-B., Liu, C., Qiao, J.-X., Yu, X.-Q., Wang, Y.-F., 2019. *Wolbachia*-induced expression of kenny gene in testes affects male fertility in *Drosophila melanogaster*. Insect Sci. https://doi.org/10.1111/1744-7917.12730

Bonneau, M., Atyame, C., Beji, M., Justy, F., Cohen-Gonsaud, M., Sicard, M., Weill, M., 2018. *Culex pipiens* crossing type diversity is governed by an amplified and polymorphic operon of *Wolbachia*. Nat. Commun. 9. https://doi.org/10.1038/s41467-017-02749-w

Bordenstein, S.R., O'Hara, F.P., Werren, J.H., 2001. *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. Nature 409, 707–710. https://doi.org/10.1038/35055543

Brennan, L.J., Haukedal, J.A., Earle, J.C., Keddie, B., Harris, H.L., 2012. Disruption of redox homeostasis leads to oxidative DNA damage in spermatocytes of *Wolbachia*-infected *Drosophila simulans*. Insect Mol. Biol. 21, 510–520. https://doi.org/10.1111/j.1365-2583.2012.01155.x

Brennan, L.J., Keddie, B.A., Braig, H.R., Harris, H.L., 2008. The endosymbiont *Wolbachia pipientis* induces the expression of host antioxidant proteins in an *Aedes albopictus* cell line. PloS One 3, e2083. https://doi.org/10.1371/journal.pone.0002083

Brucker, R.M., Bordenstein, S.R., 2012. Speciation by symbiosis. Trends Ecol. Evol. 27, 443–451. https://doi.org/10.1016/j.tree.2012.03.011

Cai, Z., Yan, L.-J., 2013. Protein Oxidative Modifications: Beneficial Roles in Disease and Health. J. Biochem. Pharmacol. Res. 1, 15–26.

Caragata, E.P., Dutra, H.L.C., Moreira, L.A., 2016. Inhibition of Zika virus by *Wolbachia* in *Aedes aegypti*. Microb. Cell 3, 293–295. https://doi.org/10.15698/mic2016.07.513

Casiraghi, M., Bordenstein, S.R., Baldo, L., Lo, N., Beninati, T., Wernegreen, J.J., Werren, J.H., Bandi, C., 2005. Phylogeny of *Wolbachia pipientis* based on gltA, groEL and ftsZ gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the *Wolbachia* tree. Microbiol.-Sgm 151, 4015–4022. https://doi.org/10.1099/mic.0.28313-0

Charlesworth, J., Weinert, L.A., Araujo, E.V., Welch, J.J., 2019. *Wolbachia*, *Cardinium* and climate: an analysis of global data. Biol. Lett. 15, 20190273. https://doi.org/10.1098/rsbl.2019.0273

610 Chen, D., Zheng, W., Lin, A., Uyhazi, K., Zhao, H., Lin, H., 2012. Pumilio 1 Suppresses Multiple Activators
611 of p53 to Safeguard Spermatogenesis. Curr. Biol. 22, 420–425.
612 https://doi.org/10.1016/j.cub.2012.01.039
613 Chen, H., Ronau, J.A., Beckmann, J., Hochstrasser, M., 2019. A *Wolbachia* nuclease and its binding
614 partner provide a distinct mechanism for cytoplasmic incompatibility. Proc. Natl. Acad. Sci. 116,
615 22314–22321. https://doi.org/10.1073/pnas.1914571116
616 Cho, K.-O., Kim, G.-W., Lee, O.-K., 2011. *Wolbachia* bacteria reside in host Golgi-related vesicles whose
617 position is regulated by polarity proteins. PloS One 6, e22703.
618 https://doi.org/10.1371/journal.pone.0022703
619 Cunningham, B.C., Wells, J.A., 1989. High-resolution epitope mapping of hGH-receptor interactions by
620 alanine-scanning mutagenesis. Science 244, 1081–1085.
621 https://doi.org/10.1126/science.2471267
622 Driver, C., Georgiou, A., Georgiou, G., 2004. The contribution by mitochondrially induced oxidative
623 damage to aging in *Drosophila melanogaster*. Biogerontology 5, 185–192.
624 https://doi.org/10.1023/B:BGEN.0000031156.75376.e3
625 Duffy, J.B., 2002. GAL4 system in *Drosophila*: a fly geneticist's Swiss army knife. Genes. N. Y. N 2000 34,
626 1–15. https://doi.org/10.1002/gene.10150
627 Fattouh, N., Cazevieille, C., Landmann, F., 2019. *Wolbachia* endosymbionts subvert the endoplasmic
628 reticulum to acquire host membranes without triggering ER stress. PLoS Negl. Trop. Dis. 13,
629 e0007218. https://doi.org/10.1371/journal.pntd.0007218
630 Ferree, P.M., Sullivan, W., 2006. A genetic test of the role of the maternal pronucleus in *Wolbachia*-
631 induced cytoplasmic incompatibility in *Drosophila melanogaster*. Genetics 173, 839–847.
632 https://doi.org/10.1534/genetics.105.053272
633 Forbes, A., Lehmann, R., 1998. Nanos and Pumilio have critical roles in the development and function of
634 *Drosophila* germline stem cells. Dev. Camb. Engl. 125, 679–690.
635 Ghate, N.B., Kim, J., Shin, Y., Situ, A., Ulmer, T.S., An, W., 2019. p32 is a negative regulator of p53
636 tetramerization and transactivation. Mol. Oncol. 13, 1976–1992. https://doi.org/10.1002/1878-
637 0261.12543
638 Govind, S., 2008. Innate immunity in *Drosophila*: Pathogens and pathways. Insect Sci. Online 15, 29–43.
639 https://doi.org/10.1111/j.1744-7917.2008.00185.x
640 Guy, B., Krell, T., Sanchez, V., Kennel, A., Manin, C., Sodoyer, R., 2005. Do Th1 or Th2 sequence motifs
641 exist in proteins?: Identification of amphipatic immunomodulatory domains in *Helicobacter*
642 *pylori* catalase. Immunol. Lett. 96, 261–275. https://doi.org/10.1016/j.imlet.2004.09.011
643 Hilgenboecker, K., Hammerstein, P., Schlattmann, P., Telschow, A., Werren, J.H., 2008. How many
644 species are infected with *Wolbachia*? - a statistical analysis of current data. Fems Microbiol. Lett.
645 281, 215–220. https://doi.org/10.1111/j.1574-6968.2008.01110.x
646 Homa, S.T., Vessey, W., Perez-Miranda, A., Riyait, T., Agarwal, A., 2015. Reactive Oxygen Species (ROS) in
647 human semen: determination of a reference range. J. Assist. Reprod. Genet. 32, 757–764.
648 https://doi.org/10.1007/s10815-015-0454-x
649 Horowitz, H., Berg, C.A., 1996. The *Drosophila* pipsqueak gene encodes a nuclear BTB-domain-containing
650 protein required early in oogenesis. Dev. Camb. Engl. 122, 1859–1871.
651 Jaenike, J., Dyer, K.A., Cornish, C., Minhas, M.S., 2006. Asymmetrical reinforcement and *Wolbachia*
652 infection in *Drosophila*. PLoS Biol. 4, e325. https://doi.org/10.1371/journal.pbio.0040325
653 Kittayapong, P., Kaeothaisong, N.-O., Ninphanomchai, S., Limohpasmanee, W., 2018. Combined sterile
654 insect technique and incompatible insect technique: sex separation and quality of sterile *Aedes*
655 *aegypti* male mosquitoes released in a pilot population suppression trial in Thailand. Parasit.
656 Vectors 11, 657. https://doi.org/10.1186/s13071-018-3214-9

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

Kumar, G.A., Subramaniam, K., 2018. PUF-8 facilitates homologous chromosome pairing by promoting proteasome activity during meiotic entry in *C. elegans*. Development dev.163949. https://doi.org/10.1242/dev.163949

Layton, E.M., On, J., Perlmutter, J.I., Bordenstein, S.R., Shropshire, J.D., 2019. Paternal grandmother age affects the strength of *Wolbachia*-induced cytoplasmic incompatibility in *Drosophila melanogaster*. mBio 10. https://doi.org/10.1128/mBio.01879-19

LePage, D., Bordenstein, S.R., 2013. *Wolbachia*: Can we save lives with a great pandemic? Trends Parasitol. 29, 385–393. https://doi.org/10.1016/j.pt.2013.06.003

LePage, D.P., Metcalf, J.A., Bordenstein, Sarah R., On, J., Perlmutter, J.I., Shropshire, J.D., Layton, E.M., Funkhouser-Jones, L.J., Beckmann, J., Bordenstein, Seth R., 2017. Prophage WO genes recapitulate and enhance *Wolbachi*a-induced cytoplasmic incompatibility. Nature 543, 243–247. https://doi.org/10.1038/nature21391

Lindsey, A., Rice, D.W., Bordenstein, Sarah R., Brooks, A.W., Bordenstein, Seth R., Newton, I.L.G., 2018. Evolutionary genetics of cytoplasmic incompatibility genes *cifA* and *cifB* in prophage WO of *Wolbachia*. Genome Biol. Evol. 10, 434–451. https://doi.org/10/gcvmkm

Loew, O., 1900. A new enzyme of general occurrence in organisms. Science 11, 701–702. https://doi.org/10.1126/science.11.279.701

Macdonald, P.M., 1992. The *Drosophila* pumilio gene: an unusually long transcription unit and an unusual protein. Dev. Camb. Engl. 114, 221–232.

Mi, D., Ou, X., Li, P., Peng, G., Liu, Y., Guo, R., Mu, Z., Li, F., Holmes, K., Qian, Z., 2019. Glycine 29 Is Critical for Conformational Changes of the Spike Glycoprotein of Mouse Hepatitis Virus A59 Triggered by either Receptor Binding or High pH. J. Virol. 93. https://doi.org/10.1128/JVI.01046-19

Moreira, L.A., Iturbe-Ormaetxe, I., Jeffery, J.A., Lu, G., Pyke, A.T., Hedges, L.M., Rocha, B.C., Hall-Mendelin, S., Day, A., Riegler, M., Hugo, L.E., Johnson, K.N., Kay, B.H., McGraw, E.A., van den Hurk, A.F., Ryan, P.A., O'Neill, S.L., 2009. A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, Chikungunya, and Plasmodium. Cell 139, 1268–1278. https://doi.org/10.1016/j.cell.2009.11.042

Nishanth, M.J., Simon, B., 2020. Functions, mechanisms and regulation of Pumilio/Puf family RNA binding proteins: a comprehensive review. Mol. Biol. Rep. 47, 785–807. https://doi.org/10.1007/s11033-019-05142-6

O'Connor, L., Plichart, C., Sang, A.C., Brelsfoard, C.L., Bossin, H.C., Dobson, S.L., 2012. Open release of male mosquitoes infected with a *Wolbachia* biopesticide: field performance and infection containment. PLoS Negl. Trop. Dis. 6, e1797. https://doi.org/10.1371/journal.pntd.0001797

O'Neill, S.L., 2018. The use of *Wolbachia* by the World Mosquito Program to interrupt transmission of *Aedes aegypti* transmitted viruses. Adv. Exp. Med. Biol. 1062, 355–360. https://doi.org/10.1007/978-981-10-8727-1_24

Parisi, M., Lin, H., 1999. The *Drosophila* pumilio gene encodes two functional protein isoforms that play multiple roles in germline development, gonadogenesis, oogenesis and embryogenesis. Genetics 153, 235–250.

Pinto, S.B., Stainton, K., Harris, S., Kambris, Z., Sutton, E.R., Bonsall, M.B., Parkhill, J., Sinkins, S.P., 2013. Transcriptional regulation of *Culex pipiens* mosquitoes by *Wolbachia* influences cytoplasmic incompatibility. PLoS Pathog. 9, e1003647. https://doi.org/10.1371/journal.ppat.1003647

703 Rasgon, J.L., 2008. Using Predictive Models to Optimize *Wolbachia*-Based Strategies for Vector-Borne
704 Disease Control, in: Aksoy, S. (Ed.), Transgenesis and the Management of Vector-Borne Disease.
705 Springer New York, New York, NY, pp. 114–125. https://doi.org/10.1007/978-0-387-78225-6_10
706 Reynolds, K.T., Hoffmann, A.A., 2002. Male age, host effects and the weak expression or non-expression
707 of cytoplasmic incompatibility in *Drosophila* strains infected by maternally transmitted
708 *Wolbachia*. Genet. Res. 80, 79–87.
709 Serbus, L.R., Casper-Lindley, C., Landmann, F., Sullivan, W., 2008. The genetics and cell biology of
710 *Wolbachia*-host interactions. Annu. Rev. Genet. 42, 683–707.
711 https://doi.org/10.1146/annurev.genet.41.110306.130354
712 Shropshire, J.D., Bordenstein, S.R., 2019. Two-by-one model of cytoplasmic incompatibility: synthetic
713 recapitulation by transgenic expression of *cifA* and *cifB* in *Drosophila*. PLOS Genet. 15,
714 e1008221. https://doi.org/10.1371/journal.pgen.1008221
715 Shropshire, J.D., Bordenstein, S.R., 2016. Speciation by symbiosis: the microbiome and behavior. mBio 7,
716 e01785-15. https://doi.org/10.1128/mBio.01785-15
717 Shropshire, J.D., Leigh, B., Bordenstein, Sarah R., Duplouy, A., Riegler, M., Brownlie, J.C., Bordenstein,
718 Seth R., 2019. Models and nomenclature for cytoplasmic incompatibility: caution over
719 premature conclusions – a response to Beckmann et al. Trends Genet. 0.
720 https://doi.org/10.1016/j.tig.2019.03.004
721 Shropshire, J.D., On, J., Layton, E.M., Zhou, H., Bordenstein, S.R., 2018. One prophage WO gene rescues
722 cytoplasmic incompatibility in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. 115, 4987–
723 4991. https://doi.org/10.1073/pnas.1800650115
724 Siegmund, T., Lehmann, M., 2002. The *Drosophila* Pipsqueak protein defines a new family of helix-turn-
725 helix DNA-binding proteins. Dev. Genes Evol. 212, 152–157. https://doi.org/10.1007/s00427-
726 002-0219-2
727 Taylor, M.J., Bordenstein, S.R., Slatko, B., 2018. Microbe Profile: *Wolbachia*: a sex selector, a viral
728 protector and a target to treat filarial nematodes. Microbiology 164, 1345–1347.
729 https://doi.org/10.1099/mic.0.000724
730 Tram, U., Sullivan, W., 2002. Role of delayed nuclear envelope breakdown and mitosis in *Wolbachia*-
731 induced cytoplasmic incompatibility. Science 296, 1124–1126.
732 https://doi.org/10.1126/science.1070536
733 Turelli, M., 1994. Evolution of incompatibility-inducing microbes and their hosts. Evol. Int. J. Org. Evol.
734 48, 1500–1513. https://doi.org/10.1111/j.1558-5646.1994.tb02192.x
735 Weidmann, C.A., Goldstrohm, A.C., 2012. *Drosophila* Pumilio Protein Contains Multiple Autonomous
736 Repression Domains That Regulate mRNAs Independently of Nanos and Brain Tumor. Mol. Cell.
737 Biol. 32, 527–540. https://doi.org/10.1128/MCB.06052-11
738 Weinert, L.A., Araujo-Jnr, E.V., Ahmed, M.Z., Welch, J.J., 2015. The incidence of bacterial endosymbionts
739 in terrestrial arthropods. Proc R Soc B 282, 20150249. https://doi.org/10.1098/rspb.2015.0249
740 Wong Sak Hoi, J., Dumas, B., 2010. Ste12 and Ste12-Like Proteins, Fungal Transcription Factors
741 Regulating Development and Pathogenicity. Eukaryot. Cell 9, 480–485.
742 https://doi.org/10.1128/EC.00333-09
743 Yamada, R., Floate, K.D., Riegler, M., O'Neill, S.L., 2007. Male development time influences the strength
744 of *Wolbachia*-induced cytoplasmic incompatibility expression in *Drosophila melanogaster*.
745 Genetics 177, 801–808. https://doi.org/10.1534/genetics.106.068486
746 Yang, H., Ahmad, Z.A., Song, Y., 2020. Molecular insight for the role of key residues of calreticulin in its
747 binding activities: A computational study. Comput. Biol. Chem. 85, 107228.
748 https://doi.org/10.1016/j.compbiolchem.2020.107228
749 Yu, B., Huang, Z., 2015. Variations in Antioxidant Genes and Male Infertility. BioMed Res. Int. 2015.
750 https://doi.org/10.1155/2015/513196

751 Zhang, Y., 2009. I-TASSER: fully automated protein structure prediction in CASP8. Proteins 77 Suppl 9,
752 100–113. https://doi.org/10.1002/prot.22588
753 Zheng, X., Zhang, D., Li, Y., Yang, C., Wu, Y., Liang, X., Liang, Y., Pan, X., Hu, L., Sun, Q., Wang, X., Wei, Y.,
754 Zhu, J., Qian, W., Yan, Z., Parker, A.G., Gilles, J.R.L., Bourtzis, K., Bouyer, J., Tang, M., Zheng, B.,
755 Yu, J., Liu, J., Zhuang, J., Hu, Zhigang, Zhang, M., Gong, J.-T., Hong, X.-Y., Zhang, Z., Lin, L., Liu, Q.,
756 Hu, Zhiyong, Wu, Z., Baton, L.A., Hoffmann, A.A., Xi, Z., 2019. Incompatible and sterile insect
757 techniques combined eliminate mosquitoes. Nature. https://doi.org/10.1038/s41586-019-1407-
758 9
759 Zheng, Y., Shen, W., Bi, J., Chen, M.-Y., Wang, R.-F., Ai, H., Wang, Y.-F., 2019. Small RNA analysis provides
760 new insights into cytoplasmic incompatibility in *Drosophila melanogaster* induced by *Wolbachia*.
761 J. Insect Physiol. 118, 103938. https://doi.org/10.1016/j.jinsphys.2019.103938
762 Zug, R., Hammerstein, P., 2015. *Wolbachia* and the insect immune system: what reactive oxygen species
763 can tell us about the mechanisms of *Wolbachia*–host interactions. Front. Microbiol. 6.
764 https://doi.org/10.3389/fmicb.2015.01201
765 Zug, R., Hammerstein, P., 2012. Still a host of hosts for *Wolbachia*: analysis of recent data suggests that
766 40% of terrestrial arthropod species are infected. PLOS ONE 7, e38544.
767 https://doi.org/10.1371/journal.pone.0038544
768

769

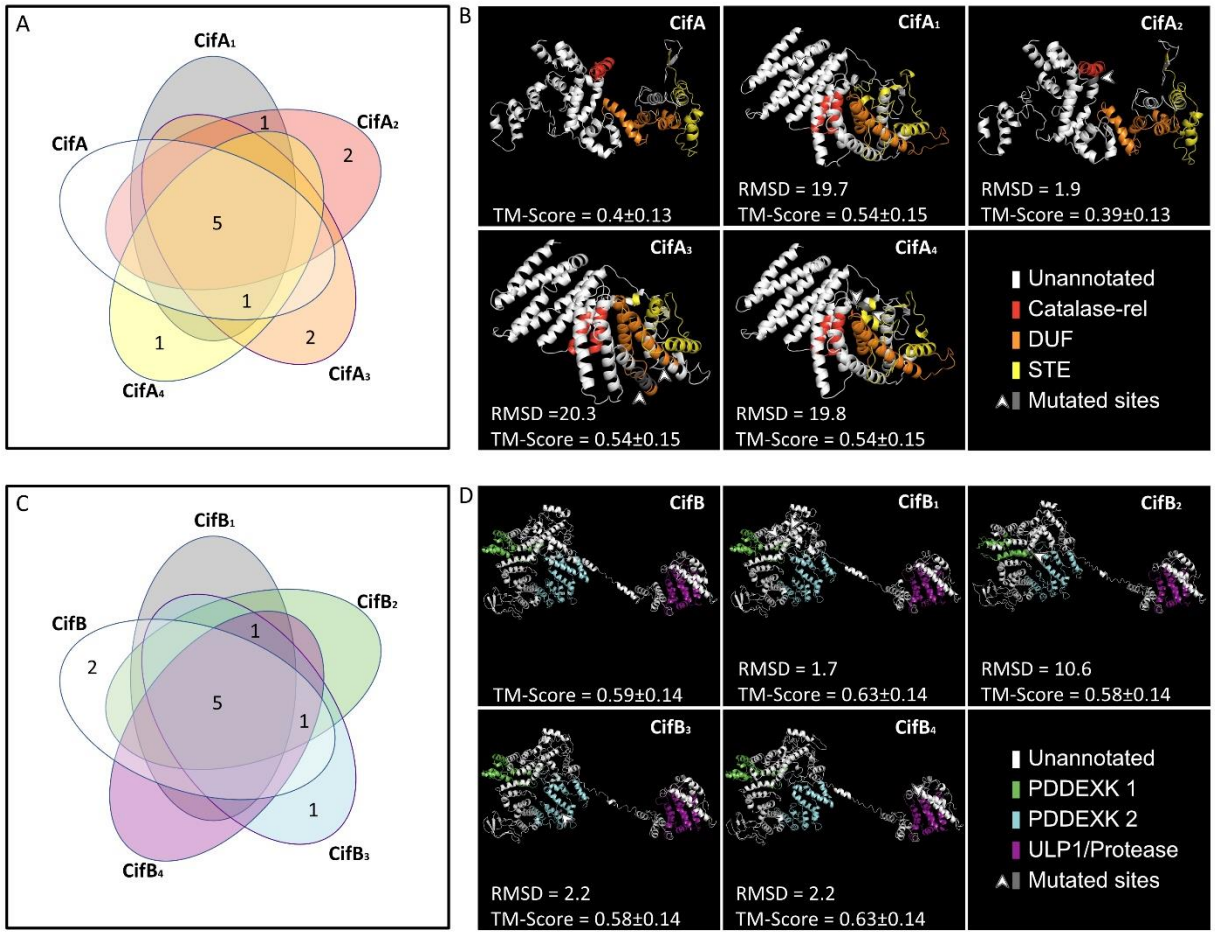

Figure S1. Summary of I-TASSER structural predictions for CifA, CifB, and their mutants. (A, C) Venn-diagrams showing the number of PDB hits shared between wild-type and mutant (A) CifA and (C) CifB proteins. (B, D) I-TASSER uses these PDB hits to generate structural predictions for (C) CifA and (D) CifB. TM-scores range from 0-1 where 1 is the highest confidence. RMSD scores are from pairwise alignments of mutant proteins with the wild-type in PyMol. Higher RMSD scores represent more distance between the superimposed proteins. Mutated sites in the tertiary structure are indicated with a white arrow. Domain annotations were based on previous sequence analyses (Lindsey et al., 2018). Details regarding the PDB hits are reported in Table S2.

Table S1. P values associated with all statistical comparisons made in main and extended data hatch rate and cytology figures. M=male, F=female ,+=Wolbachia infected, −=Wolbachia uninfected.

Table S2. Protein structural prediction software I-TASSER identifies homologous protein domains found in all of our Cif homologs, and those that differ.

Table S3. Nucleotide and protein sequences used in this study.

Table S4. Primers used for genotyping and sanger sequencing if cif transgenes.

Source Data File 1. All raw data from main figures, supplemental figures, and replicate data experiments in this study.