LETTER

Prophage WO genes recapitulate and enhance *Wolbachia*–induced cytoplasmic incompatibility

Daniel P. LePage[1]*, Jason A. Metcalf[1]*, Sarah R. Bordenstein[1], Jungmin On[1], Jessamyn I. Perlmutter[1], J. Dylan Shropshire[1], Emily M. Layton[1], Lisa J. Funkhouser-Jones[1], John F. Beckmann[2] & Seth R. Bordenstein[1,3]

The genus *Wolbachia* is an archetype of maternally inherited intracellular bacteria that infect the germline of numerous invertebrate species worldwide. They can selfishly alter arthropod sex ratios and reproductive strategies to increase the proportion of the infected matriline in the population. The most common reproductive manipulation is cytoplasmic incompatibility, which results in embryonic lethality in crosses between infected males and uninfected females. Females infected with the same *Wolbachia* strain rescue this lethality. Despite more than 40 years of research[1] and relevance to symbiont-induced speciation[2,3], as well as control of arbovirus vectors[4–6] and agricultural pests[7], the bacterial genes underlying cytoplasmic incompatibility remain unknown. Here we use comparative and transgenic approaches to demonstrate that two differentially transcribed, co-diverging genes in the eukaryotic association module of prophage WO[8] from *Wolbachia* strain wMel recapitulate and enhance cytoplasmic incompatibility. Dual expression in transgenic, uninfected males of *Drosophila melanogaster* crossed to uninfected females causes embryonic lethality. Each gene additively augments embryonic lethality in crosses between infected males and uninfected females. Lethality associates with embryonic defects that parallel those of wild-type cytoplasmic incompatibility and is notably rescued by wMel-infected embryos in all cases. The discovery of cytoplasmic incompatibility factor genes *cifA* and *cifB* pioneers genetic studies of prophage WO-induced reproductive manipulations and informs the continuing use of *Wolbachia* to control dengue and Zika virus transmission to humans.

We hypothesized that the genes responsible for cytoplasmic incompatibility (CI) (Extended Data Fig. 1a) are present in all CI-inducing *Wolbachia* strains and absent or divergent in non-CI strains; we also predicted that these genes are expressed in the gonads of infected insects. To elucidate CI candidates, we determined the core genome shared by the CI-inducing *Wolbachia* strains wMel (from *D. melanogaster*), wRi (from *Drosophila simulans*), wPip (Pel strain from *Culex pipiens*), and wRec (from *Drosophila recens*), while excluding the pan-genome of the mutualistic strain wBm (from *Brugia malayi*). This yielded 113 gene families representing 161 unique wMel genes (Fig. 1a and Supplementary Table 1). We further narrowed this list by comparing it with (1) homologues of genes previously determined by comparative genomic hybridization to be absent or divergent in the strain wAu[9], a non-CI strain, (2) homologues to genes highly expressed at the RNA level in wVitA-infected *Nasonia vitripennis* ovaries, and (3) homologues detected at the protein level in wPip (Buckeye)-infected *C. pipiens* ovaries. We included ovarian data with the reasoning that CI genes might be generally expressed in infected reproductive tissues, or that the CI induction and rescue genes might be the same. Remarkably, only two genes, wMel locus tags WD0631 and WD0632, were shared among all four gene subsets (Fig. 1b and

Supplementary Tables 2–4). Notably, the homologue of WD0631 in the *Wolbachia* strain wPip was found at the protein level in the fertilized spermathecae of infected mosquitoes, lending support to its role in reproductive manipulation[10].

We analysed the evolution and predicted protein domains of these two genes and found that their homologues are always paired within the eukaryotic association module of prophage WO[8], and they co-diverged into three distinct phylogenetic groups that we designate types I, II, and III (Fig. 1c, e and Supplementary Table 5). These relationships are not evident in the phylogeny of the *Wolbachia* cell division gene *ftsZ*, which exhibits the typical bifurcation of supergroup A and B *Wolbachia* (Extended Data Fig. 1b), or in the phylogeny of prophage WO baseplate assembly gene *gpW* (Extended Data Fig. 1c). This suggests that homologues of WD0631 and WD0632 evolve under different evolutionary pressures than genes in the core *Wolbachia* genome or in a structural module of phage WO.

Type I variants are the most prevalent among ten sequenced *Wolbachia* strains, and are always associated with large prophage WO regions that often lack tail genes (Extended Data Fig. 2); it is unclear whether these WO regions forge fully intact or defective interfering particles. The functions of type I WD0631 homologues are unknown, although type I WD0632 homologues contain weak homology to a putative Peptidase_C48 domain (wMel, National Center for Biotechnology Information (NCBI) conserved domain $E = 6.69 \times 10^{-4}$, Fig. 1f), a key feature of Ulp1 (ubiquitin-like-specific protease) proteases[10]. Type II variants are located within more complete phage haplotypes (Extended Data Fig. 2), but the WD0632 homologues are truncated and lack recognized protein domains (Fig. 1f). Notably, all *Wolbachia* strains that contain type II variants contain at least one pair of intact type I variants. Type III variants possess WD0631 homologues with a weakly predicted cytochrome C552 domain involved in nitrate reduction (wNo, NCBI conserved domain $E = 3.79 \times 10^{-3}$), while type III WD0632 homologues contain weak homology to the PD-(D/E)XK nuclease superfamily (wNo, NCBI conserved domain $E = 1.15 \times 10^{-3}$) and to a transmembrane domain predicted by the transmembrane hidden Markov model[11] (Fig. 1f). Finally, a putative type IV variant encoding a carboxy (C)-terminal PD-(D/E)XK nuclease superfamily (NCBI conserved domain $E = 3.69 \times 10^{-3}$) was identified in *Wolbachia* strains wPip and wAlbB, but not included in phylogenetic analyses because the WD0632 homologues are highly divergent (28% identity across 17% of the protein) and do not appear in reciprocal BLASTp analyses. The predicted functions of type III and IV protein domains are not well understood, but a homologue of the putative nuclease domain was previously found in a selfish genetic element that mediates embryonic lethality in *Tribolium* beetles[12]. Uncertain annotations and substantial unknown sequence across all of the phylogenetic types necessitate caution in extrapolating definitive gene functions. Importantly, the region of shared homology among the WD0632 homologues (Fig. 1f) is outside the putative

[1]Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37235, USA. [2]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA. [3]Department of Pathology, Microbiology, and Immunology, Vanderbilt University, Nashville, Tennessee 37235, USA.
*These authors contributed equally to this work.
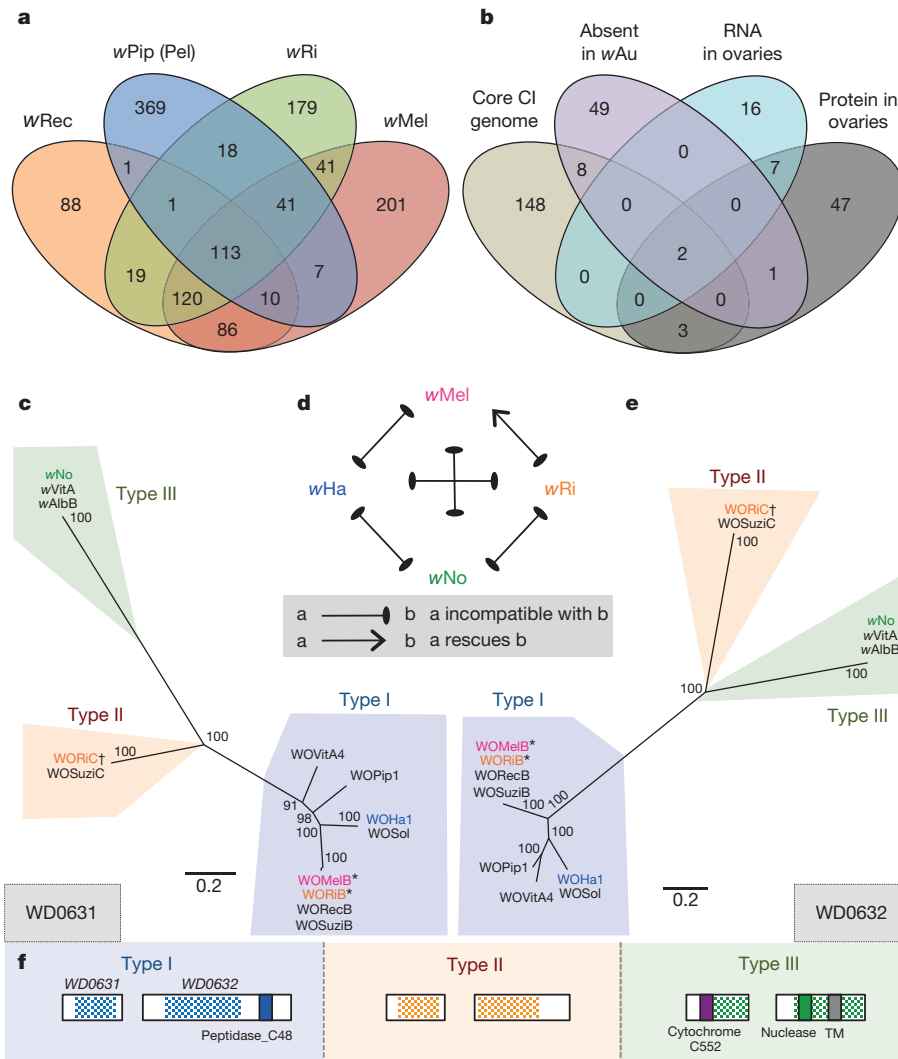
Figure 1 | Comparative analyses reveal WD0631 and WD0632 in the eukaryotic association module of prophage WO as candidate CI genes. a, Venn diagram illustrating the number of unique and shared gene families from four CI-inducing *Wolbachia* strains. **b**, Venn diagram illustrating the number of unique and shared *w*Mel genes matching each criteria combination. **c, e**, Bayesian phylogenies of (**c**) WD0631 and (**e**) WD0632 and their homologues, on the basis of a core 256-amino-acid (aa) alignment of WD0631 reciprocal BLASTp hits and a core 462-aa alignment of WD0632 reciprocal BLASTp hits. When multiple similar copies exist in the same strain, only one copy is shown. Consensus support values are shown at the nodes. Both trees are based on the JTT+G model of evolution and are unrooted. **d**, CI patterns correlate with WD0631/ WD0632 sequence homology. *w*Ri rescues *w*Mel and both share a similar set of homologues (*). The inability of *w*Mel to rescue *w*Ri correlates with a type (†) that is present in *w*Ri but absent in *w*Mel. Likewise, bidirectional incompatibility of all other crosses correlates to divergent homologues. This diagram was adapted from ref. 30. **f**, Protein architecture of the WD0631 and WD0632 types is conserved for each clade and is classified according to the WD0632-like domain. TM, transmembrane. Dotted shading represents the region of shared homology used to construct phylogenetic trees. For **c** and **e**, the WO-prefix indicates a specific phage WO haplotype and the *w*-prefix refers to a 'WO-like island', a small subset of conserved phage genes, within that specific *Wolbachia* strain.

C-terminal Peptidase_C48 domain, suggesting that the unannotated regions represent an ancestral CI sequence core that warrants closer inspection.

Consistent with a role in CI, the degree of relatedness and presence/ absence of homologues of WD0631 and WD0632 between *Wolbachia* strains correlates with known patterns of bidirectional incompatibility (Fig. 1d). Among the strains *w*Ri, *w*Ha, and *w*No, only *w*Ri rescues *w*Mel-induced CI in same-species crosses[13,14]. We postulate that this is due to *w*Ri and *w*Mel sharing highly related type I homologues (99% amino-acid identity), and thus probably sharing a rescue factor, while *w*Ri also has a type II homologue that may explain its ability to induce CI against *w*Mel. Meanwhile, bidirectionally incompatible pairs are highly divergent, with only 29–68% amino-acid identity (Extended Data Fig. 3a). Additionally, variation in CI strength between strains appears to correlate with the number of copies of the WD0631/WD0632 pair (Extended Data Fig. 3b). Strains with only one copy, such as *w*Mel, have a comparatively weak CI phenotype, while those with two or three copies, such as *w*Ri and *w*Ha, cause strong CI[14].

Given the various lines of evidence that associate these two genes with CI, we next examined the functional role of WD0631 and WD0632 in CI. For comparison, the following control genes were also used: WD0034, which has a predicted PAZ (Piwi, Argonaut, and Zwille) domain (NCBI conserved domain $E = 1.85 \times 10^{-18}$); WD0508, a prophage gene annotated as a putative transcriptional regulator with two helix–turn–helix domains (NCBI conserved domain $E = 9.29 \times 10^{-12}$) in the Octomom region; and WD0625, a prophage gene annotated as a DUF2466 with a JAB1/MPN/Mov34 metalloenzyme (JAMM) domain (NCBI conserved domain $E = 1.60 \times 10^{-41}$). We first examined the expression of these genes in the testes of *w*Mel-infected, 1-day-old and 7-day-old *D. melanogaster* males. Since CI strength decreases significantly in aged males[15], we predicted that a
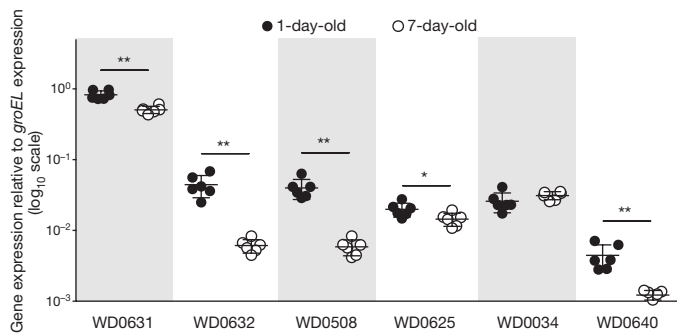
Figure 2 | Relative expression of CI candidate and prophage WO genes decreases as males age. RNA expression in 1-day-old versus 7-day-old testes, normalized to expression of *groEL* in *w*Mel-infected *D. melanogaster* testes from the fastest-developing males. Values denote $2^{-\Delta C_t}$. $n = 6$ independent pools of 20 testes for each group. Bars, mean ± s.d. *$P < 0.05$, **$P < 0.01$ by Mann–Whitney *U*-test. This experiment was performed once. Exact *P* values are provided in Supplementary Table 7.

CI factor would be expressed at a lower level in 7-day-old males versus 1-day-old males that both emerged on day 1 of the cross. Indeed, WD0631 and WD0632 showed a significantly lower transcription level in aged males (Fig. 2). Moreover, WD0631 exhibited 18.6- and 83.0-fold higher expression than WD0632 for young and aged males, respectively (Fig. 2). Coupled with RNA-seq expression data[16] and operon predictor algorithms, evidence suggests that these genes are not generally acting as an operon in *w*Mel. Both prophage-associated control genes, WD0508 and WD0625, also exhibited this age-dependent expression pattern, but the non-prophage gene WD0034 did not (Fig. 2). WD0640, which encodes prophage WO structural protein GpW, was also reduced in older males, suggesting that prophage genes in general are relatively downregulated in 7-day-old testes (Fig. 2). The phenomenon of decreased CI in older males was not due to decreases in *Wolbachia* titre over time, as the copy number of *Wolbachia groEL* relative to *D. melanogaster* *rp49* increased as males aged, and there was no significant difference in absolute *Wolbachia* gene copies between 1-day-old and 7-day-old males (Extended Data Fig. 4a, b). Since CI expression is also correlated with male development time, we examined gene expression in early emerging 'older brothers' (emerged on day 1) and later emerging 'younger brothers' (emerged on day 5). Expression was statistically equivalent for WD0631 (Extended Data Fig. 4c), and slightly reduced in younger brothers for WD0632 (Extended Data Fig. 4d). These results are consistent with a small younger brother effect[17], although we did not observe a statistically significant effect on CI penetrance (Extended Data Fig. 4e).

To directly test involvement of these genes in CI, we generated transgenic *D. melanogaster* expressing genes using an upstream activating sequence (UAS), since *Wolbachia* itself cannot be genetically transformed. We used a *nanos*-Gal4 driver line for tissue-specific expression in the male and female germline[18,19]. We assessed CI by measuring the percentage of embryos that hatched into larvae. While wild-type (WT) CI between infected males and uninfected females led to significantly reduced hatch rates, expressing each of four candidate transgenes in uninfected (fastest-developing, 1 day old) males did not affect hatch rates when crossed to uninfected females (Fig. 3a and Extended Data Fig. 5a). These transgenes also had no effect on sex ratios (Extended Data Figs 5b and 6a). There were no phenotypic effects despite confirmed expression of each transgene in the testes (Extended Data Fig. 7a–d).

As WD0631 and WD0632 are adjacent, coevolving genes, we reasoned that dual expression of WD0631 and WD0632 might be required to induce CI. Indeed, expression of both transgenes in the same males significantly reduced hatch rates by 68% compared with uninfected WT crosses (Fig. 3b), with no effect on sex ratios (Extended Data Fig. 6b). Roughly half of the crosses with transgenic males yielded hatch rates within the range observed in WT CI (3.8 ± 5.6% hatch rate).
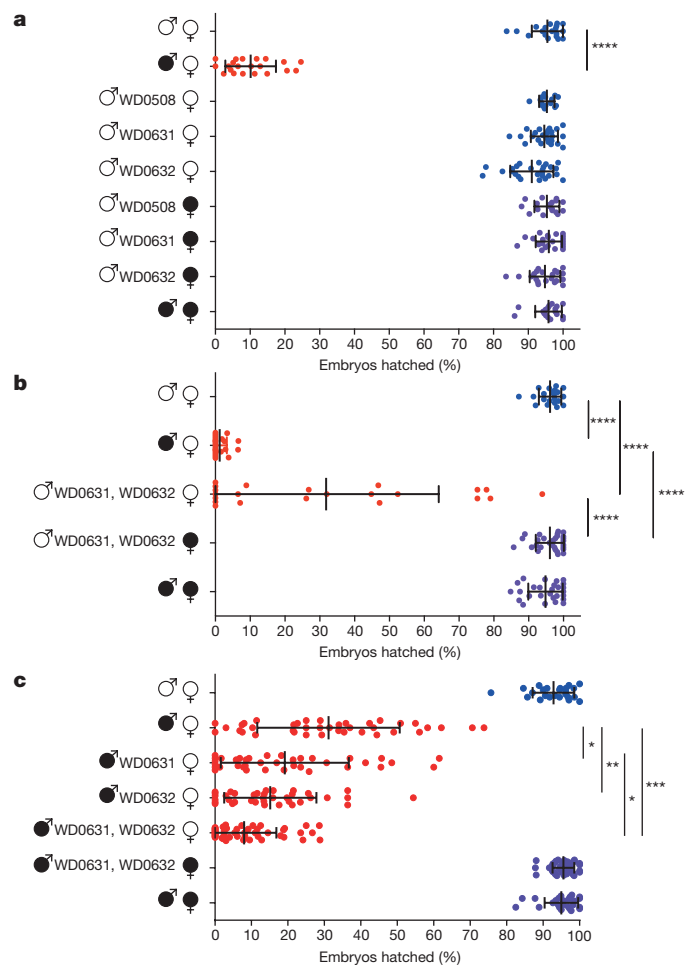
Figure 3 | Dual expression of WD0631 (*cifA*) and WD0632 (*cifB*) is necessary to induce CI-like defects. a–c, Hatch rate assays used the fastest developing males that were aged either (**a**, **b**) 1 day or (**c**) 2–4 days in parental crosses; older males express incomplete CI. Parental infection status is designated with filled symbols for *w*Mel-infected parents or open symbols for uninfected parents. Transgenic flies are labelled with their transgene to the right of their male/female symbol. Unlabelled symbols represent WT flies. Data points are coloured according to the type of cross: blue, no CI; red, CI crosses; purple, rescue crosses with *w*Mel-infected females. $n = 24$–54 for each group. Bars, mean ± s.d. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$ by analysis of variance (ANOVA) with Kruskal–Wallis test and Dunn's multiple test correction. Statistical comparisons are between all groups (**a**, **b**); or between CI crosses only (**c**). All experiments were performed at least twice, except for the increase of WT CI by WD0631, which was done once. Exact *P* values are provided in Supplementary Table 7.

Interestingly, there was a strong positive correlation between hatch rate and clutch size when both transgenes were expressed ($r_s = 0.7$; $P = 0.0003$), but not in WT CI, suggesting that dilution of transgene products across larger clutches may explain variation in transgene-induced CI. It is also possible that full transgene induction of CI requires other factors, or that transgenes are not expressed at the ideal time or place for complete CI, although transgene expression in adult testes was confirmed (Extended Data Fig. 7c, d).

Importantly, transgene-induced lethality is fully rescued in embryos of *w*Mel-infected females (Fig. 3b), indicating that these genes produce probable CI factors rather than artefacts that reduce hatch rates through off target effects that would not be rescued. We therefore name and hereafter refer to these genes as cytoplasmic incompatibility factor A (*cifA*) and B (*cifB*) for WD0631 and WD0632, respectively. Type II, III, and IV homologues are designated *cif-like* until experimental evidence demonstrates that they recapitulate CI.

To test whether *cifA* (WD0631) and *cifB* (WD0632) transgenes act additively with *Wolbachia* to enhance WT CI levels, *w*Mel-infected male flies expressing either transgene were aged 2–4 days to lower WT CI penetrance before crossing with uninfected females. In support of transgene-induced enhancement of CI, hatch rates in these aged males decreased significantly compared with WT CI crosses of the same age (Fig. 3c), with no effect on sex ratios (Extended Data Fig. 6c). In this context, wherein aged flies cause a weaker level of WT CI, the transgenes appear to add to the quantity of CI effectors in *w*Mel-infected tissues, causing stronger CI overall. This effect was not observed when control transgenes WD0508 or WD0625 were expressed individually in *w*Mel-infected males (Extended Data Fig. 8a, b). Moreover, dual expression of *cifA* and *cifB* in *w*Mel-infected flies reduced hatch rates further than either gene alone, yet was still fully rescued in embryos of *w*Mel-infected females (Fig. 3c). Adding WD0625 to *cifB* in aged *w*Mel-infected males did not increase CI beyond *cifB* alone (Extended Data Fig. 8b), and had no effect on embryonic hatch rates from crosses with 1-day-old uninfected males (Extended Data Fig. 8c). Finally, none of these gene combinations affected offspring sex ratios (Extended Data Fig. 9). Taken together, these findings support the central conclusion that *cifA* and *cifB* are both necessary to induce the CI phenotype, and they do not represent an artefact of the transgenic system.

To rule out the possibility that transgene-induced enhancement of CI in infected lines is due to increased *Wolbachia* titres, we quantitated amplicons of single-copy genes from *Wolbachia* and *D. melanogaster*. Although there were some differences in *Wolbachia* titres between infected transgenic lines (Extended Data Fig. 10a–c), the variation did not correlate with induction or magnitude of CI, signifying that decreased offspring viability was due to the direct effect of transgene products rather than changes in *Wolbachia* density. Most notably, densities significantly increased in infected flies expressing the control Octomom transgene WD0508 (Extended Data Fig. 10a) but did not enhance CI (Extended Data Fig. 8a).

Next, we tested whether transgene-induced CI associates with canonical cytological defects observed in *Wolbachia*-induced CI. Although CI is typically thought to cause failure of the first mitotic division[20,21], nearly half of the embryonic arrest in incompatible crosses occurs during advanced developmental stages in *D. simulans*[22,23], *Aedes polynesiensis*[24], and *C. pipiens*[25]. We examined embryos from control and experimental crosses after 1–2 h of development and binned their cytology into one of six phenotypes. While a few embryos in each cross were unfertilized (Fig. 4a), most embryos in compatible crosses were either in normal late-stage preblastoderm (Fig. 4b) or syncytial blastoderm stages[26] (Fig. 4c). In WT CI, significantly more embryos exhibited three defects: arrest of cellular division after two to three mitotic divisions (Fig. 4d), later stage arrest associated with chromatin bridging, as is classically associated with strong CI in *D. simulans*[21] (Fig. 4e), or arrest associated with regional failure of division in one embryo region (Fig. 4f). After blindly scoring embryo cytology, we determined that aberrant phenotypes (a, d, e, and f) were significantly more common in the offspring of dual *cifA*/*cifB* transgenic males mated to uninfected females, but these abnormalities were rescued in embryos from *w*Mel-infected females (Fig. 4g). These effects were not seen with control transgene WD0508 or with singular expression of *cifA* or *cifB* (Fig. 4h). These data again validate that transgene-induced CI, as measured through cytological defects, recapitulates WT CI. Most of the embryos arrest after two to three mitotic divisions.

This study identifies, for the first time, two differentially transcribed and codiverging prophage WO genes that recapitulate and enhance CI. These rapidly evolving genes are not chromosomal *Wolbachia* genes per se, but rather occur widely in the eukaryotic association module of prophage WO[8]. This module notably contains genes with amino-acid sequences homologous to eukaryotes or annotated to interact with animal cells, although *cifA* and *cifB* do not appear to have eukaryotic homology. CI can therefore be categorized as a prophage WO-induced
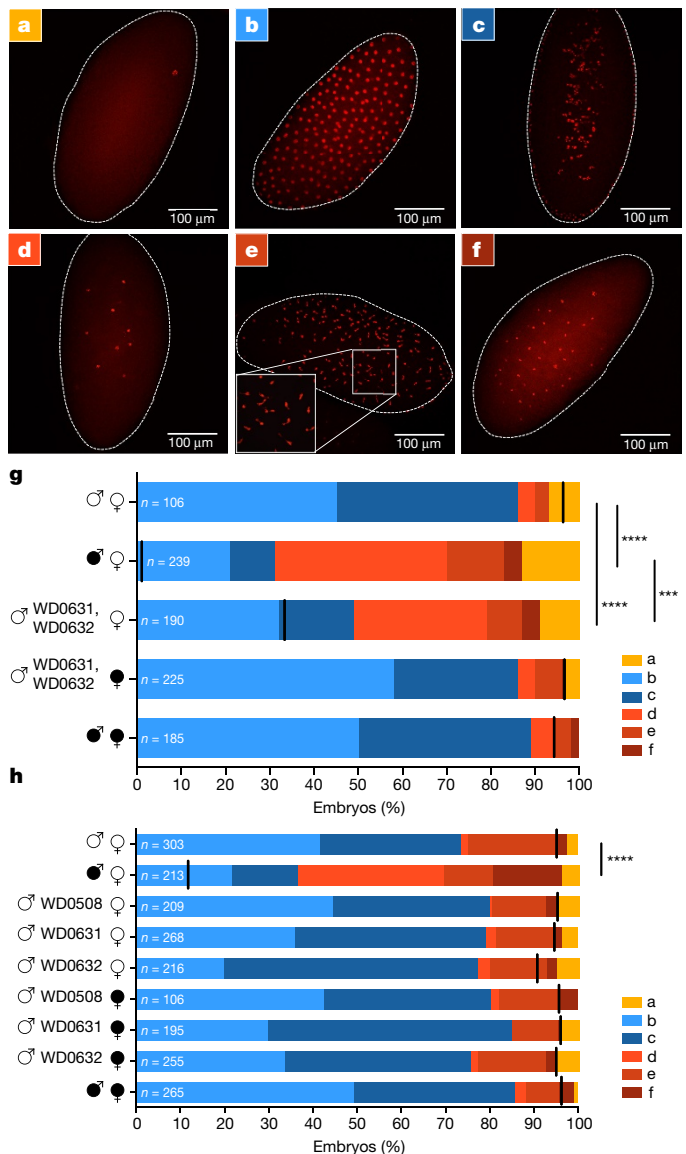
Figure 4 | Dual expression of WD0631 (*cifA*) and WD0632 (*cifB*) recapitulates CI-associated embryonic defects. a–f, Representative embryo cytology is shown for (**a**) unfertilized embryos, (**b**) normal multi-nucleated embryos at 1 h of development, (**c**) normal embryos near 2 h of development in which nuclei begin to migrate to the periphery of the cytoplasm, and three different mitotic abnormalities: (**d**) failure of nuclear division after two to three mitoses, (**e**) chromatin bridging, and (**f**) regional mitotic failure. **g**, The number of embryos with each cytological phenotype resulting from the indicated crosses is shown. Infection status is designated with filled symbols for *w*Mel-infected parents or open symbols for uninfected parents. Transgenic flies are labelled with their transgene to the right of their male/female symbol. Unlabelled symbols represent WT flies. Black lines on each graph indicates mean hatch rate for the cross. ***$P < 0.001$, ****$P < 0.0001$ by two-tailed Fisher's exact test comparing normal (phenotypes b and c) with abnormal (phenotypes a, d–f) for each cross. **h**, Quantitation of cytological defects in control crosses. Cytology for **g** was performed twice and for **h** once. Exact P values are provided in Supplementary Table 7.

phenotype rather than a *Wolbachia*-induced phenotype. We name the genes and close homologues cytoplasmic incompatibility factors A and B for WD0631 and WD0632, respectively. The *cif* name is conservatively grounded in phenotype and makes no assumptions about mechanism, which is notable because there are unannotated core regions throughout the *cif* genes that may have as much bearing on mechanism as the annotated domains.

The discovery of *cifA* and *cifB* genes that functionally recapitulate and enhance CI is the first inroad in solving the genetic basis of reproductive parasitism, a phenomenon induced worldwide in potentially millions of arthropod species[27]. These prophage WO genes have implications for microbe-assisted speciation, because they can underlie CI-induced hybrid lethality observed between closely related species of *Nasonia* and *Drosophila*[28,29]. Finally, *cifA* and *cifB* are important for arthropod pest and vector control strategies, as they could be an alternative or adjunct to current *Wolbachia*-based efforts aimed at controlling agricultural pests or curbing arthropod-borne transmission of infectious diseases[4–7].

1. Yen, J. H. & Barr, A. R. New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens* L. *Nature* **232,** 657–658 (1971).
2. Brucker, R. M. & Bordenstein, S. R. Speciation by symbiosis. *Trends Ecol. Evol.* **27,** 443–451 (2012).
3. Shropshire, J. D. & Bordenstein, S. R. Speciation by symbiosis: the microbiome and behavior. *MBio* **7,** e01785–15 (2016).
4. O'Connor, L. *et al.* Open release of male mosquitoes infected with a *Wolbachia* biopesticide: field performance and infection containment. *PLoS Negl. Trop. Dis.* **6,** e1797 (2012).
5. Walker, T. *et al.* The *w*Mel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. *Nature* **476,** 450–453 (2011).
6. Carneiro Dutra, H. L. *et al. Wolbachia* blocks currently circulating Zika virus isolates in Brazilian *Aedes aegypti* mosquitoes. *Cell Host Microbe* **19,** 771–774 (2016).
7. Zabalou, S. *et al. Wolbachia*-induced cytoplasmic incompatibility as a means for insect pest population control. *Proc. Natl Acad. Sci. USA* **101,** 15042–15045 (2004).
8. Bordenstein, S. R. & Bordenstein, S. R. Eukaryotic association module in phage WO genomes from *Wolbachia. Nature Commun.* **7,** 13155 (2016).
9. Ishmael, N. *et al.* Extensive genomic diversity of closely related *Wolbachia* strains. *Microbiology* **155,** 2211–2222 (2009).
10. Beckmann, J. F. & Fallon, A. M. Detection of the *Wolbachia* protein WPIP0282 in mosquito spermathecae: implications for cytoplasmic incompatibility. *Insect Biochem. Mol. Biol.* **43,** 867–878 (2013).
11. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305,** 567–580 (2001).
12. Lorenzen, M. D. *et al.* The maternal-effect, selfish genetic element Medea is associated with a composite Tc1 transposon. *Proc. Natl Acad. Sci. USA* **105,** 10085–10089 (2008).
13. Zabalou, S. *et al.* Multiple rescue factors within a *Wolbachia* strain. *Genetics* **178,** 2145–2160 (2008).
14. Poinsot, D., Bourtzis, K., Markakis, G., Savakis, C. & Mercot, H. *Wolbachia* transfer from *Drosophila melanogaster* into *D. simulans*: host effect and cytoplasmic incompatibility relationships. *Genetics* **150,** 227–237 (1998).
15. Reynolds, K. T. & Hoffmann, A. A. Male age, host effects and the weak expression or non-expression of cytoplasmic incompatibility in *Drosophila* strains infected by maternally transmitted *Wolbachia. Genet. Res.* **80,** 79–87 (2002).
16. Gutzwiller, F. *et al.* Dynamics of *Wolbachia pipientis* gene expression across the *Drosophila melanogaster* life cycle. *G3* **5,** 2843–2856 (2015).
17. Yamada, R., Floate, K. D., Riegler, M. & O'Neill, S. L. Male development time influences the strength of *Wolbachia*-induced cytoplasmic incompatibility expression in *Drosophila melanogaster. Genetics* **177,** 801–808 (2007).
18. Rørth, P. Gal4 in the *Drosophila* female germline. *Mech. Dev.* **78,** 113–118 (1998).
19. White-Cooper, H. Tissue, cell type and stage-specific ectopic gene expression and RNAi induction in the *Drosophila* testis. *Spermatogenesis* **2,** 11–22 (2012).
20. Serbus, L. R., Casper-Lindley, C., Landmann, F. & Sullivan, W. The genetics and cell biology of *Wolbachia*-host interactions. *Annu. Rev. Genet.* **42,** 683–707 (2008).
21. Landmann, F., Orsi, G. A., Loppin, B. & Sullivan, W. *Wolbachia*-mediated cytoplasmic incompatibility is associated with impaired histone deposition in the male pronucleus. *PLoS Pathog.* **5,** e1000343 (2009).
22. Lassy, C. W. & Karr, T. L. Cytological analysis of fertilization and early embryonic development in incompatible crosses of *Drosophila simulans. Mech. Dev.* **57,** 47–58 (1996).
23. Callaini, G., Riparbelli, M. G., Giordano, R. & Dallai, R. Mitotic defects associated with cytoplasmic incompatibility in *Drosophila simulans. J. Invertebr. Pathol.* **67,** 55–64 (1996).
24. Wright, J. D. & Barr, A. R. *Wolbachia* and the normal and incompatible eggs of *Aedes polynesiensis* (Diptera: Culicidae). *J. Invertebr. Pathol.* **38,** 409–418 (1981).
25. Duron, O. & Weill, M. *Wolbachia* infection influences the development of *Culex pipiens* embryo in incompatible crosses. *Heredity* **96,** 493–500 (2006).
26. Foe, V. E., Odell, G. M. & Edgar, B. A. in *The Development of* Drosophila melanogaster (eds Bate, M. & Martinez-Arias, A.) Ch. 3 (Cold Spring Harbor Laboratory Press, 1993).
27. Zug, R. & Hammerstein, P. Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS ONE* **7,** e38544 (2012).
28. Jaenike, J., Dyer, K. A., Cornish, C. & Minhas, M. S. Asymmetrical reinforcement and *Wolbachia* infection in *Drosophila. PLoS Biol.* **4,** e325 (2006).
29. Bordenstein, S. R., O'Hara, F. P. & Werren, J. H. *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia. Nature* **409,** 707–710 (2001).
30. Bossan, B., Koehncke, A. & Hammerstein, P. A new model and method for understanding *Wolbachia*-induced cytoplasmic incompatibility. *PLoS ONE* **6,** e19757 (2011).

Author Contributions D.P.L. performed gene expression and hatch rate assays, embryo cytology, and assayed for transgene and infection status of flies. J.A.M. performed comparative genomics analyses, generated transgenic flies, and drafted the manuscript. Sarah R.B. performed evolutionary and bioinformatic analyses. J.O. performed hatch rates, assayed sex ratios, collected flies for all experiments, and assayed for transgene and infection status of flies. J.I.P. conducted younger brother effect experiments and performed embryo cytology. J.D.S. performed hatch rate assays, collected flies for parallel embryo cytology, and assayed for transgene and infection status of flies. E.M.L. collected flies and performed hatch rate assays. L.J.F.-J. obtained the *w*VitA transcriptome. J.F.B. obtained the *w*Pip proteome. Seth R.B. supervised the work and contributed to all experimental designs, data analysis, and data interpretation. All authors participated in manuscript preparation, editing, and final approval.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Seth R.B. (s.bordenstein@vanderbilt.edu).

Reviewer Information *Nature* thanks S. L. O'Neill, W. Sullivan and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Comparative genomics and transcriptomics. MicroScope[31] was used to select the set of genes comprising the core genomes of CI-inducing *Wolbachia* strains *w*Mel (NC_002978.6)[32], *w*Ri (NC_012416.1)[33], *w*Pip (Pel) (NC_010981.1)[34], and the recently sequenced *w*Rec (GCA_000742435.1)[35], while excluding the pan-genome of the mutualistic strain *w*Bm (NC_006833.1)[36], using cutoffs of 50% amino-acid identity and 80% alignment coverage. For the 'absent in *w*Au' criterion, *w*Au microarray data were obtained from the original authors[9] and genes that were present in CI-inducing strains *w*Ri and *w*Sim but absent or divergent in the non-CI strain *w*Au were selected.

For ovarian transcriptomics, 1-day-old females from *w*VitA-infected *N. vitripennis* 12.1 were hosted as virgins on *Sarcophaga bullata* pupae[37] for 48 h to stimulate feeding and oogenesis. Females were then dissected in RNase-free 1× PBS buffer, and their ovaries were immediately transferred to RNase-free microcentrifuge tubes in liquid nitrogen. Fifty ovaries were pooled for each of three biological replicates. Ovaries were manually homogenized with RNase-free pestles, and their RNA was extracted using the RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol for purification of total RNA from animal tissues. After RNA purification, samples were treated with RQ1 RNase-free DNase (Promega), and ethanol precipitation was performed. PCR of RNA samples with *Nasonia* primers NvS6KQTF4 and NVS6KQTR4 (ref. 38) confirmed that all samples were free of DNA contamination. RNA concentrations were measured with a Qubit 2.0 Fluorometer (Life Technologies) using the RNA HS Assay kit (Life Technologies), and approximately 5 µg of total RNA from each sample was used as input for the MICROB*Enrich* Kit (Ambion) to enrich for *Wolbachia* RNA in the samples. Bacterial-enriched RNA was then ethanol-precipitated, and rRNA was depleted from the samples using the Ribo-Zero Magnetic kit (Illumina) according to the manufacturer's protocol. Approximately 1.5 µg of enriched, rRNA-depleted RNA for each replicate was shipped to the University of Rochester Genomics Research Center for sequencing. Library preparation was performed using the Illumina ScriptSeq version 2 RNA-Seq Library Preparation kit, and all samples were run multiplexed on a single lane of the Illumina HiSeq2500 (single-end, 100 base pair reads). Raw sequence reads were trimmed and mapped to the *w*VitA genome (PRJNA213627) in CLC Genomics Workbench 8.5.1 using a minimum length fraction of 0.9, a minimum similarity fraction of 0.8, and allowing one gene hit per read. With all three replicates combined, a total of 364,765 reads out of 41,894,651 (0.87%) mapped to the *w*VitA genome, with the remaining reads mapping to the *N. vitripennis* host genome (GCF_000002325.3). All *Wolbachia* genes with at least five RNA-seq reads, with the exception of the 16S and 23S RNA genes, were selected. For non-*w*Mel data sets, the closest homologues in *w*Mel were found using BLASTp in Geneious Pro version 5.5.6 (ref. 39).

Protein extraction and mass spectrometry. Protein was extracted from *C. pipiens* tissues as described previously[10]. Ovaries from 30 *w*Pip (Buckeye)-infected mosquitoes were dissected in 100% ethanol and collected in a 1.5 ml tube filled with 100% ethanol. Pooled tissues were sonicated at 40 mA for 10 s in a Kontes GE 70.1 ultrasonic processor, and trichloroacetic acid was added to a final concentration of 10% (v/v). After centrifugation at 13,000 r.p.m. in a microcentrifuge, pellets were washed with acetone:water (9:1), dried, and stored at −20 °C. Samples were directly submitted to the University of Minnesota's Center for Mass Spectrometry and Proteomics for iTRAQ (isobaric tagging for relative and absolute quantification) analysis. Proteins were sorted according to their relative abundance as determined by the number of spectra from the single most abundant peptide. Because proteins can often produce varying amounts of detectable tryptic peptides, depending upon protein size and lysine/arginine content, we counted only the single most abundant peptide for each protein. This quantification was justified by a previous report[40] showing that the two most abundant proteins are the *Wolbachia* surface protein (WSP; WP_007302328.1) and another putative membrane protein (WP0576; WP_012481859.1). Only proteins with at least three unique peptides (95% confidence) detected were reported; using this criterion the false discovery rate was zero.

Evolutionary analyses. WD0631 (NCBI accession number AAS14330.1) and WD0632 (AAS14331.1) from *w*Mel were used as queries to perform a BLASTp search of NCBI's nonredundant (nr) protein sequence database with algorithm parameters based on a word size of six and BLOSUM62 scoring matrix[41]. Homologues were selected on the basis of the satisfaction of three criteria: (1) $E = \leq 10^{-20}$, (2) query coverage greater than 60%, and (3) presence in fully sequenced *Wolbachia* and/or phage WO genomes. FtsZ and gpW proteins were identified for all representative *Wolbachia* and phage WO genomes, respectively. Protein alignments were performed using the MUSCLE plugin[42] in Geneious Pro version 8.1.7 (ref. 39); the best models of evolution, according to corrected Akaike information criteria[43], were estimated using the ProtTest server[44]; and phylogenetic trees were built using the MrBayes plugin[45] in Geneious. Putative functional domains were identified using NCBI's BLASTp, Wellcome Trust Sanger Institute's

PFAM database[46], a transmembrane hidden Markov model[11], and EMBL's Simple Modular Architecture Research Tool (SMART)[47]. WD0631/WD0632 protein homology (percentage amino-acid identity) was based on a 1:1 BLASTp analysis for each pair. Prophage/WO-like island association for each pair of genes was based on prophage regions identified in a previous study[8].

Gene expression assays and *Wolbachia* titres. For the male age effect, native expression of CI candidates was tested with RT–qPCR on replicate pools of 20 pairs of testes from the fastest-developing virgin males that were aged 1 day or 7 days. RNA was extracted with a Qiagen RNeasy mini kit, DNase treated with TURBO DNase (Life Technologies), and cDNA generated with Superscript III Reverse Transcriptase (Invitrogen). Primer sequences are listed in Supplementary Table 6. Quantitative PCR was performed on a Bio-Rad CFX-96 Real-Time System using iTaq Universal SYBR Green Supermix (Bio-Rad). Thirty cycles of PCR were performed against positive controls (extracted DNA), negative controls (water), RNA, and cDNA with the following conditions: 95 °C 2 min, 30× (95 °C 15 s, 56 °C 30 s, 72 °C 30 s), 72 °C 5 min. Values of $2^{-\Delta C_t}$ between the target gene and housekeeping gene *groEL* were used to determine relative gene expression. These experiments were performed once with multiple replicates for each condition.

For experiments on the younger brother effect, replicate pools of 20 pairs of testes were collected from the fastest-developing virgin males that emerged on the first day (older brothers) or fifth day (younger brothers). Male siblings for the younger brother effect analysis were also collected concurrently for hatch rates as described for hatch rate assays by crossing the *w*Mel-infected males to 3- to 5-day-old *w*Mel-infected or uninfected females. RNA was extracted using the Direct-zol RNA MiniPrep Kit (Zymo), DNase treated with DNA-free (Ambion, Life Technologies), cDNA was generated with SuperScript VILO (Invitrogen), and RT–qPCR was run using iTaq Universal SYBR Green (Bio-Rad). Primers, PCR conditions, and analysis were the same as for the male age effect above. These experiments were performed once with multiple replicates for each condition.

For gene expression in Extended Data Fig. 7, six pools of six pairs of testes were dissected from parents used in hatch rate assays from a repeat of Fig. 3a and Extended Data Fig. 5. In samples designated 'High CI' and 'No CI', the males correspond to crosses that had lower or normal hatch rates, respectively. For all other samples, the flies used were chosen at random. RNA was extracted using the same method as the younger brother experiment above. Thirty cycles of PCR were performed against positive controls (extracted DNA), negative controls (water), RNA, and cDNA with PCR conditions described above. Gel image size and brightness were adjusted in some cases for clarity. These experiments were performed once.

For the *Wolbachia* titres, pools of testes were dissected from 15 males in ice-cold PBS. For Extended Data Fig. 10a–c, brothers of those used in the corresponding hatch rates were used. DNA was extracted using a Gentra Puregene Tissue kit (Qiagen). qPCR was done as described above. Absolute quantification was achieved by comparing all experimental samples with a standard curve generated on the same plate. Primers are listed in Supplementary Table 6. qPCR conditions were as follows: 50 °C 10 min, 95 °C 5 min, 40× (95 °C 10 s, 55 °C 30 s), 95 °C 30 s. To obtain a more accurate *Wolbachia*:host cell ratio, it was assumed that each host cell had two copies of *rp49* and each *Wolbachia* cell had one copy of *groEL*. These experiments were performed once but with a sample size of eight for each condition.

Fly rearing. *D. melanogaster* were reared on a standard cornmeal- and molasses-based media. Stocks were maintained at 25 °C while virgin flies were stored at room temperature. During virgin collections, stocks were kept at 18 °C overnight and 25 °C during the day. All flies were kept on a 12-h light/dark cycle. *Wolbachia*-uninfected lines were generated through tetracycline treatment for three generations. Briefly, tetracycline was dissolved in ethanol and then diluted in water to a final concentration of 1 mg/ml. One millilitre of this solution was added to 50 ml of media (final concentration of 20 µg/ml). Freshly treated media was used for each generation. Infection status was confirmed with PCR using Wolb_F and Wolb_R3 primers[48], and flies were reared on untreated media for at least three additional generations to allow for mitochondrial recovery before being used[49].

Transgenic flies. Each CI candidate gene was cloned into the pTIGER plasmid for transformation and expression in *D. melanogaster*[50]. pTIGER, a pUASp-based vector designed for germline expression, exhibits targeted integration into the *D. melanogaster* genome using PhiC31 integrase[51] and tissue-specific, inducible expression through the Gal4–UAS system[52]. Cloning was performed using standard molecular biology techniques and plasmids were purified and Sanger-sequenced for confirmation before injection. At least 200 *D. melanogaster* embryos were injected per gene by Best Gene (Chino Hills, California), and transformants were selected on the basis of w^+ eye colour. All transgenic lines were made in the *yw D. melanogaster* background, and each was an isofemale line derived from the offspring of a single transformant. Homozygous lines were maintained when possible, or heterozygous flies were maintained when homozygous transgenics were inviable (WD0625/CyO). WD0508 and WD0631 insertion was performed

with the y^1 M{vas-int.Dm}ZH-2A w^*; P{CaryP}attP40 line. WD0625 was inserted into BSC9723 with the genotype y^1 M{vas-int.Dm}ZH-2A w^*; PBac{y+-attP-3B} VK00002. WD0632 insertion was done using BSC8622 with the genotype y^1 w^{67c23}; P{CaryP}attP2.

Hatch rate and sex ratio assays. Parental females were either infected or uninfected y^1w^* flies (wMel-infected or uninfected) and aged for 2–6 days before crossing. Uninfected y^1w^* flies were generated as described for transgenic lines. Parental males were created by crossing nanos-Gal4 virgin females (wMel-infected or uninfected) with either y^1w^* or UAS-candidate gene-transgenic, uninfected males. Only the first males emerging between 0 and 30 h from these crosses were used in CI assays to control for the younger-brother effect associated with CI[17]. To test whether CI can be increased by transgenes, virgin, day 1 males were aged for 2–4 days before crossing to reduce the level of WT CI. Within experiments, care was taken to match the age of males between experimental and control crosses. Thirty-two to 64 individual crosses were set up for each crossing condition. The flies used were chosen at random from the desired group on the basis of age, sex, and genotype. These sample sizes were based on previous studies of CI in *D. melanogaster* that detected significant differences between treatment groups[53].

To perform the hatch rate assays, a male and female pair was placed in an 8-ounce, round bottom, polypropylene *Drosophila* stock bottle. A grape-juice–agar plate with a small amount of yeast mix smeared on top was placed in the bottle opening and affixed with tape. To create grape-juice–agar plates, 12.5 g of agar was mixed in 350 ml of de-ionized water and autoclaved. In a separate flask, 10 ml of ethanol was used to dissolve 0.25 g tegosept (methyl 4-hyrdoxybenzoate). Welch's grape juice (150 ml) was added to the tegosept mix, combined with the agar, and poured into lids from 35 ×10-mm culture dishes (CytoOne).

Hatch rate bottles were placed in an incubator at 25 °C overnight (~16 h). After this initial incubation, the grape plates were discarded and replaced with freshly yeasted plates. After an additional 24 h, the adult flies were removed and frozen for expression analysis, and the embryos on each plate were counted. The counting was not blinded. These plates were then incubated at 25 °C for 36 h before the number of unhatched embryos was counted. Larvae from each pair of flies were moved from these plates using a probe and placed in vials of standard fly media with one vial being used for each individual grape plate to be assayed for sex ratios at adulthood. A total of 10–20 vials were used for each cross type. Any crosses with fewer than 25 embryos laid were discarded from the hatching analysis while vials with fewer than ten adults emerging were discarded from the sex ratio analysis. Statistical analysis and outlier removal, using the ROUT method, were performed using Graphpad Prism version 6 software.

Embryo cytology. Embryos were collected in a fashion similar to hatch rate assays except bottles contained 60–80 females and 15–20 males. All flies used were brothers and sisters of those used during corresponding hatch rates. Embryo collections and hatch rates were performed side-by-side. After initial mating overnight, fresh grape plates with yeast were provided and removed after 60 min. The embryo-covered plates were then placed in the incubator at 25 °C for a further 60 min to ensure each embryo was at least 1–2 h old. Embryos were then moved to a small mesh basket and dechorionated in 50% bleach for 1–3 min. These were then washed in embryo wash solution (0.7% NaCl, 0.05% Triton X-100) and moved to a small vial containing ~2 ml heptane. An equal amount of methanol was added to the vial and then vigorously shaken for 15 s. After the embryos settled, the upper heptane layer and as much methanol as possible were removed, and the embryos were moved into ~500 μl fresh methanol in a 1.5 ml microcentrifuge tube. Embryos were stored overnight at 4 °C. The old methanol was then removed and replaced with 250 μl of fresh methanol along with 750 μl of PBTA (1× PBS, 1% BSA, 0.05% Triton X-100, 0.02% sodium azide). After inverting the tube several times, the solution was removed and replaced with 500 μl PBTA. Embryos were then rehydrated for 15 min on a rotator at room temperature. After rehydrating, the PBTA was replaced with 100 μl of a 10 mg/ml solution of RNase A (Clontech Labs) and incubated at 37 °C for 2 h. The RNase was then removed and embryos were washed several times with PBS followed by a final wash with PBS–azide (1× PBS, 0.02% sodium azide). After removing the PBS–azide, embryos were mounted on glass slides with ProLong Diamond Antifade (Life Technologies) spiked with propidium iodide (Sigma-Aldrich) to a final concentration of 1 μg/ml. Imaging was performed at the Vanderbilt University Medical Center Cell Imaging Shared Resource using a Zeiss LSM 510 META inverted confocal microscope. All scores were performed blind (researcher was not aware of which slide represented which cross) and image analysis was done using ImageJ software[54]. Matched scoring, where embryos were derived from a side-by-side hatch rate, was performed once for conditions shown in Fig. 4h and twice for Fig. 4g.

Statistical analyses. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment, except scoring of cytology (Fig. 4), which was done blindly.

All statistical analyses used GraphPad Prism software (either Prism 6 or online tools). When comparing gene expression levels or *Wolbachia* titres between two sets of data, we used a two-tailed, non-parametric Mann–Whitney U-test since it does not require a normal distribution of the data. For comparisons between more than two data sets, we used a non-parametric Kruskal–Wallis one-way analysis of variance test that, if significant, was followed by a Dunn's test of multiple comparisons. This allowed robust testing between all data groups while avoiding multiple test bias. For the cytology studies, embryos were classified as either 'normal' or 'CI-like' in a 2 × 2 contingency table, and statistical differences between the groups were calculated using a Fisher's exact test.

Data availability. wVitA transcriptome data have been deposited in the Sequence Read Archive with Bioproject PRJNA319204 and BioSample SAMN04881412. wPip-infected ovarian proteome data have been deposited in the Proteome Xchange Consortium via the PRIDE[55] partner repository with the data set identifier PXD004047. All other source data are available as Supplementary Information with this publication.

31. Vallenet, D. et al. MicroScope: a platform for microbial genome annotation and comparative genomics. *Database* **2009**, bap021 (2009).
32. Wu, M. et al. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* **2**, e69 (2004).
33. Klasson, L. et al. The mosaic genome structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*. *Proc. Natl Acad. Sci. USA* **106**, 5725–5730 (2009).
34. Klasson, L. et al. Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. *Mol. Biol. Evol.* **25**, 1877–1887 (2008).
35. Metcalf, J. A., Jo, M., Bordenstein, S. R., Jaenike, J. & Bordenstein, S. R. Recent genome reduction of *Wolbachia* in *Drosophila recens* targets phage WO and narrows candidates for reproductive parasitism. *PeerJ* **2**, e529 (2014).
36. Foster, J. et al. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol.* **3**, e121 (2005).
37. Werren, J. H. & Loehlin, D. W. Rearing *Sarcophaga bullata* fly hosts for *Nasonia* (parasitoid wasp). *Cold Spring Harb. Protoc.* **2009**, pdb.prot5308 (2009).
38. Bordenstein, S. R. & Bordenstein, S. R. Temperature affects the tripartite interactions between bacteriophage WO, *Wolbachia*, and cytoplasmic incompatibility. *PLoS ONE* **6**, e29106 (2011).
39. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
40. Beckmann, J. F., Markowski, T. W., Witthuhn, B. A. & Fallon, A. M. Detection of the *Wolbachia*-encoded DNA binding protein, HU beta, in mosquito gonads. *Insect Biochem. Mol. Biol.* **43**, 272–279 (2013).
41. Johnson, M. et al. NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9 (2008).
42. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
43. Hurvich, C. M. & Tsai, C.-L. A corrected Akaike information criterion for vector autoregressive model selection. *J. Time Ser. Anal.* **14**, 271–279 (1993).
44. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104–2105 (2005).
45. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
46. Finn, R. D. et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285 (2016).
47. Letunic, I., Doerks, T. & Bork, P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* **40**, D302–D305 (2012).
48. Casiraghi, M., Anderson, T. J., Bandi, C., Bazzocchi, C. & Genchi, C. A phylogenetic analysis of filarial nematodes: comparison with the phylogeny of *Wolbachia* endosymbionts. *Parasitology* **122**, 93–103 (2001).
49. Chatzispyrou, I. A., Held, N. M., Mouchiroud, L., Auwerx, J. & Houtkooper, R. H. Tetracycline antibiotics impair mitochondrial function and its experimental use confounds research. *Cancer Res.* **75**, 4446–4449 (2015).
50. Ferguson, S. B., Blundon, M. A., Klovstad, M. S. & Schupbach, T. Modulation of gurken translation by insulin and TOR signaling in *Drosophila*. *J. Cell Sci.* **125**, 1407–1419 (2012).
51. Groth, A. C., Fish, M., Nusse, R. & Calos, M. P. Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. *Genetics* **166**, 1775–1782 (2004).
52. Southall, T. D., Elliott, D. A. & Brand, A. H. The GAL4 system: a versatile toolkit for gene expression in *Drosophila*. *Cold Spring Harb. Protoc.* **2008**, pdb.top49 (2008).
53. LePage, D. P., Jernigan, K. K. & Bordenstein, S. R. The relative importance of DNA methylation and Dnmt2-mediated epigenetic regulation on *Wolbachia* densities and cytoplasmic incompatibility. *PeerJ* **2**, e678 (2014).
54. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).
55. Vizcaíno, J. A. et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, D447–D456 (2016).
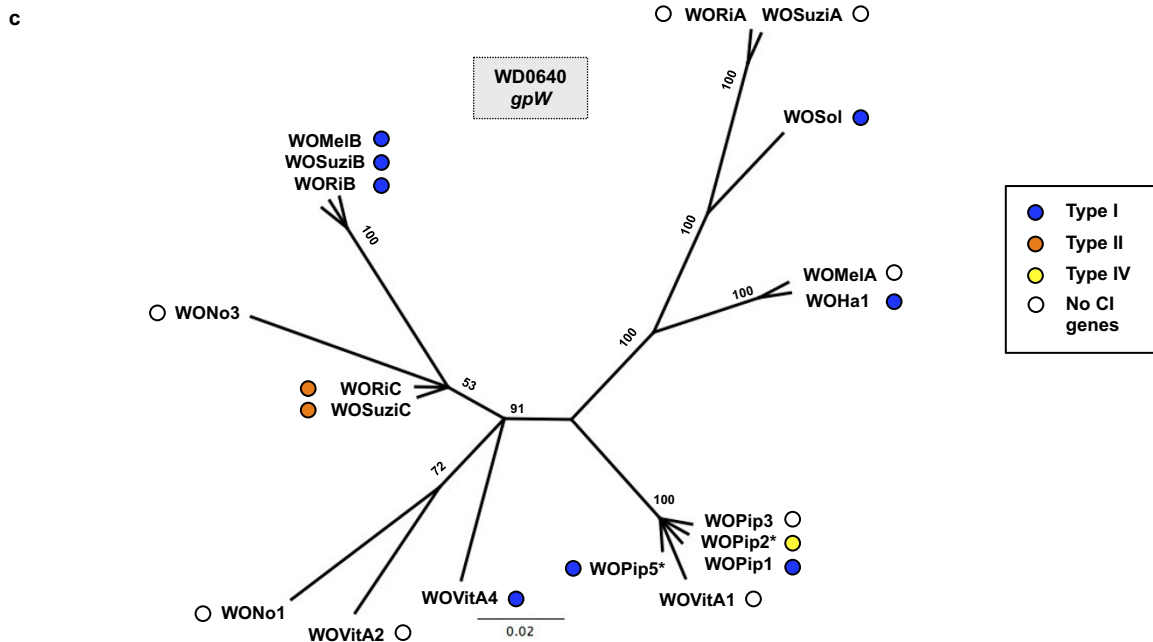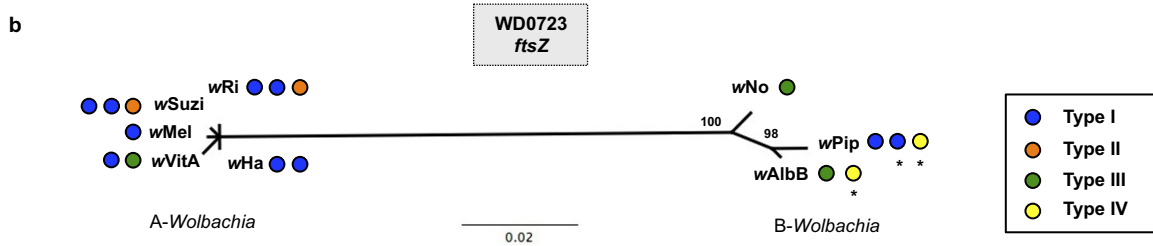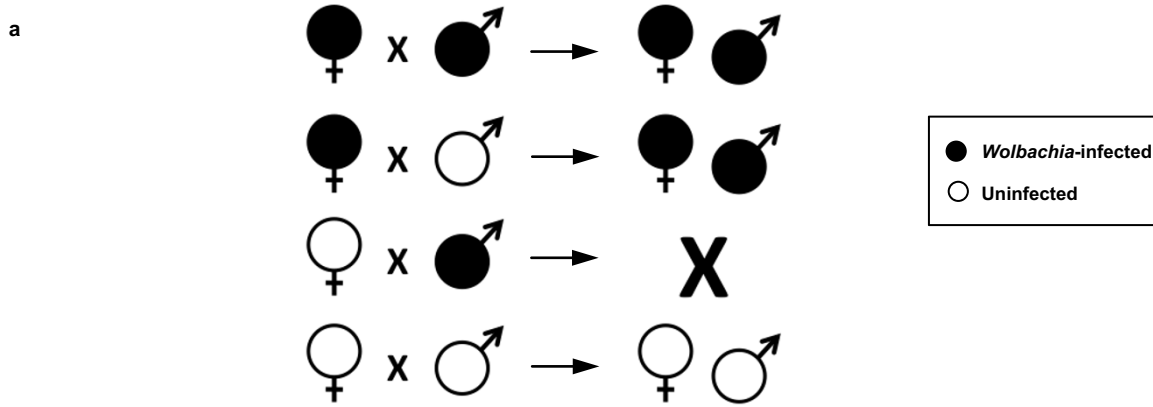
Extended Data Figure 1 | CI and the evolution of *Wolbachia* and prophage WO genes. a, The effect of parental *Wolbachia* infection on progeny viability and infection status. CI (embryonic inviability) occurs in crosses between *Wolbachia*-infected males and uninfected females. *Wolbachia*-infected females mated to infected males rescue the inviability. **b**, Bayesian phylogenies based on a 393-aa alignmen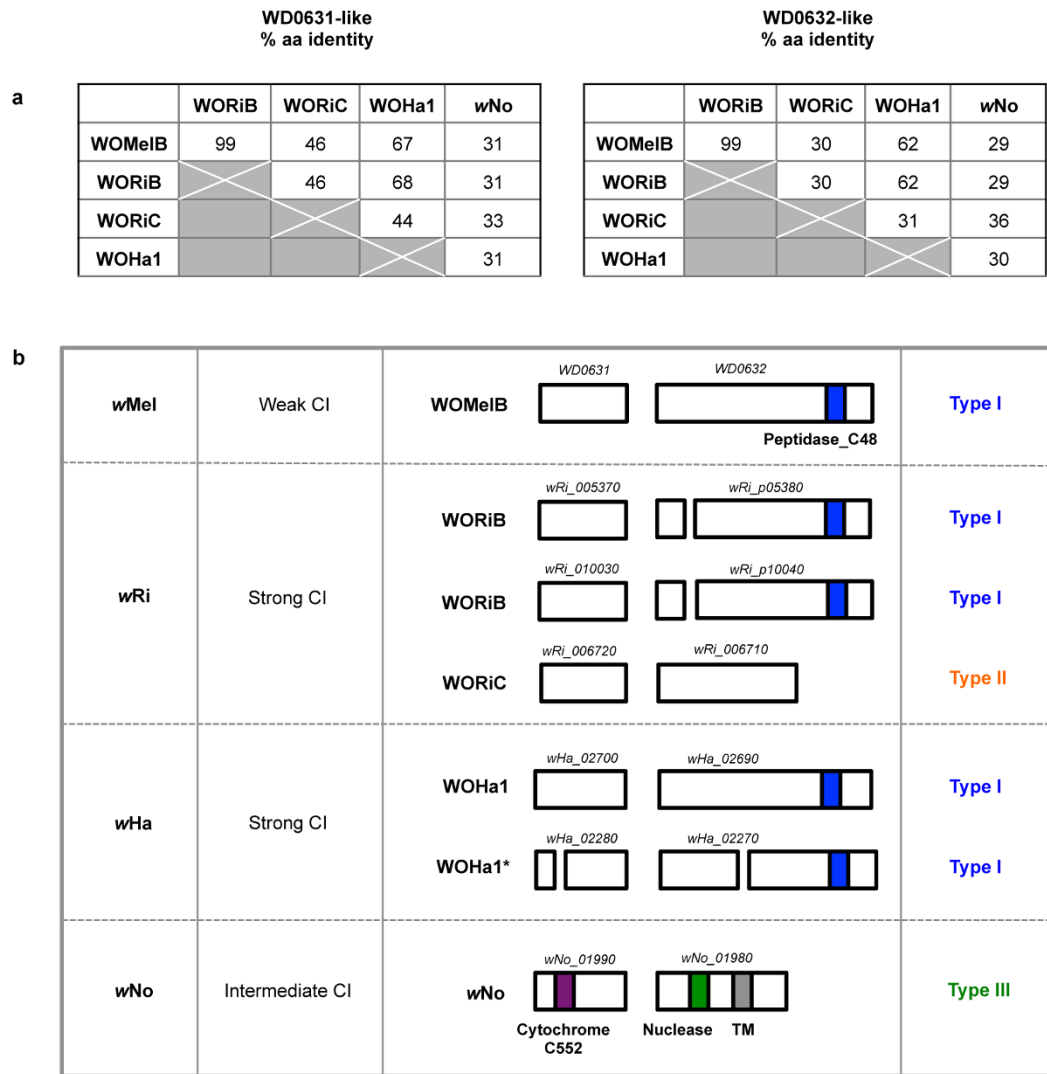t of WD0723, the *w*Mel *ftsZ* gene, and its homologues, and (**c**) a 70-aa alignment of WD0640, the phage WO *gpW* gene, and its homologues. Trees are based on JTT+G and CpRev+I models of evolution, respectively, and are unrooted. Consensus support values are shown at the nodes. Asterisk indicates that the CI genes are not included in Fig. 1. The WOPip5 homologue is truncated while the WOPip2 and second *w*AlbB homologues are highly divergent from WD0632.
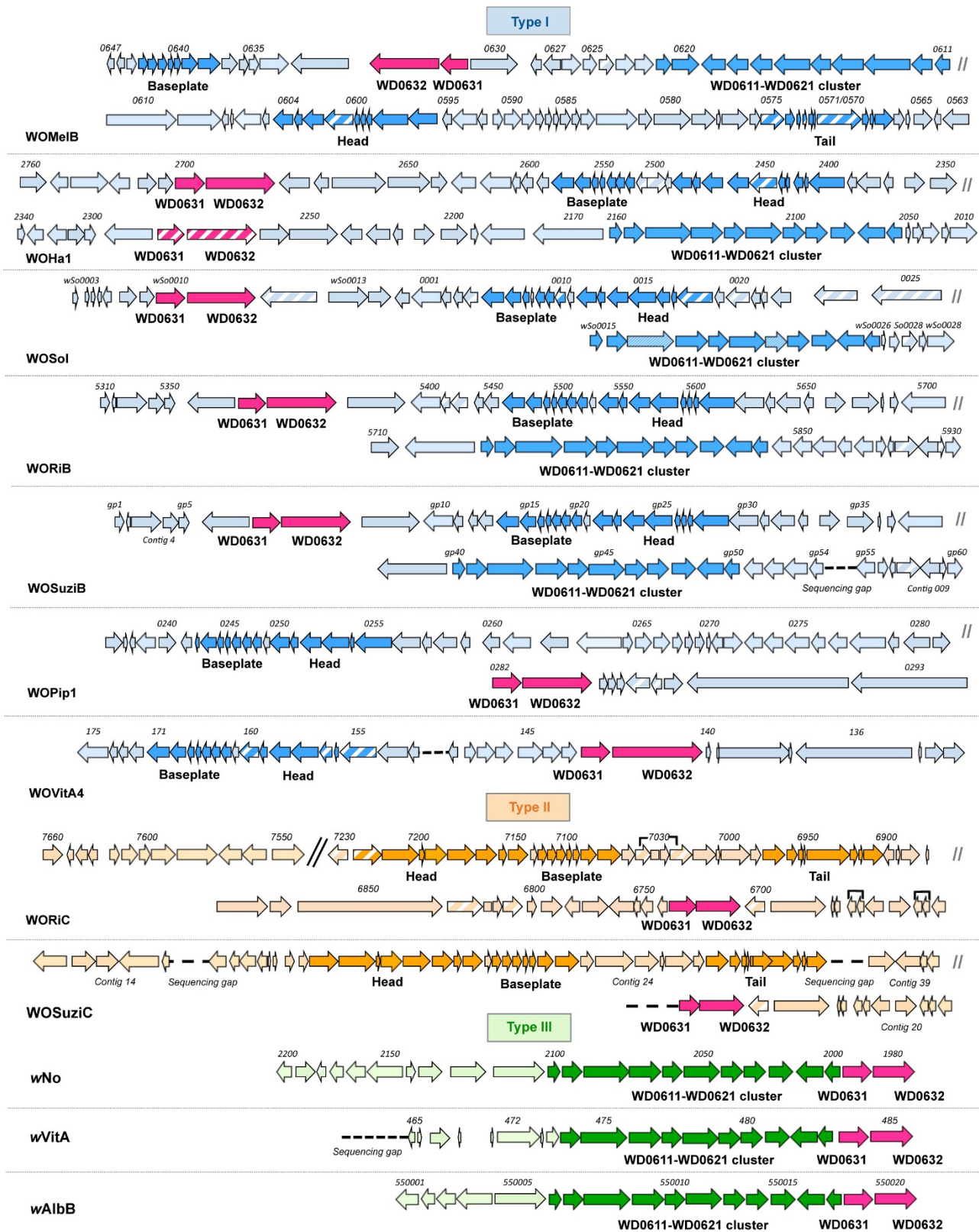
Extended Data Figure 2 | WD0631/WD0632 homologues associate with the eukaryotic association module in prophage WO regions. CI gene homologues are labelled and coloured pink. Structural modules are labelled as baseplate, head, or tail. The WD0611–WD0621 label highlights a conserved gene cluster that is often associated with the CI genes. Only one phage haplotype is shown per *Wolbachia* strain when multiple copies of the same type are present.

WD0631-like
% aa identity

	WORiB	WORiC	WOHa1	*w*No
WOMelB	99	46	67	31
WORiB		46	68	31
WORiC			44	33
WOHa1				31

WD0632-like
% aa identity

	WORiB	WORiC	WOHa1	*w*No
WOMelB	99	30	62	29
WORiB		30	62	29
WORiC			31	36
WOHa1				30

a

b

*w*Mel	Weak CI	WOMelB	*WD0631* *WD0632* Peptidase_C48	Type I
*w*Ri	Strong CI	WORiB	*wRi_005370* *wRi_p05380*	Type I
		WORiB	*wRi_010030* *wRi_p10040*	Type I
		WORiC	*wRi_006720* *wRi_006710*	Type II
*w*Ha	Strong CI	WOHa1	*wHa_02700* *wHa_02690*	Type I
		WOHa1*	*wHa_02280* *wHa_02270*	Type I
*w*No	Intermediate CI	*w*No	*wNo_01990* *wNo_01980* Cytochrome C552 Nuclease TM	Type III

Extended Data Figure 3 | *Wolbachia* CI patterns correlate with WD0631/WD0632 homologue similarity and copy number. a, The percentage aa identity between each WD0631/WD0632 homologue correlates with *Wolbachia* compatibility patterns. The only compatible cross, *w*Mel males × *w*Ri females, features close homology between WOMelB and WORiB. All other crosses are greater than 30% divergent and are bidirectionally incompatible. Each '% aa identity' value is based on the region of query coverage in a 1:1 BLASTp analysis. **b**, CI strength, protein architecture, and clade type are listed for each of the *Wolbachia* strains shown in Fig. 1d. Asterisk indicates the proteins are disrupted and not included in comparison analyses.
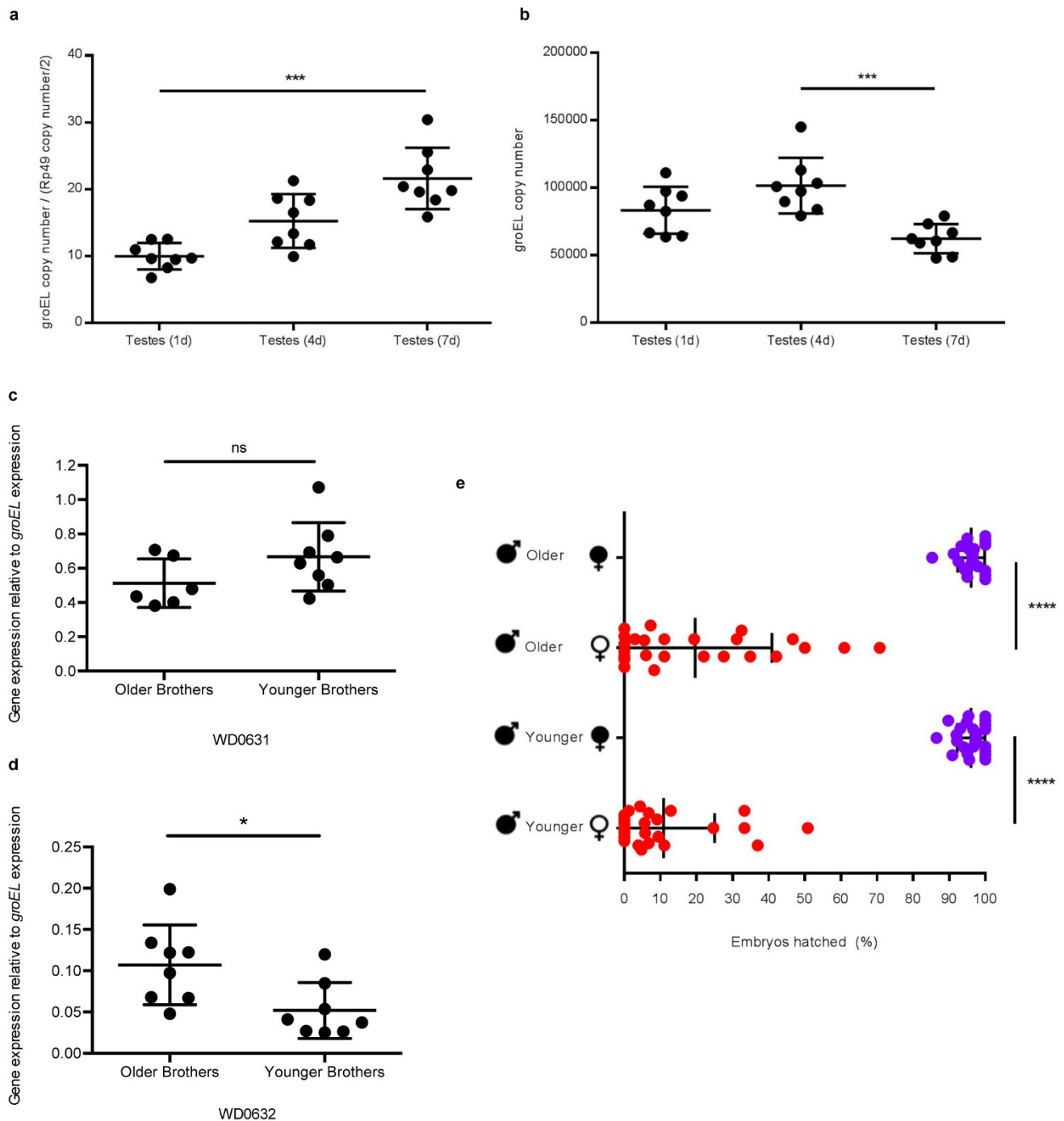
Extended Data Figure 4 | *Wolbachia* titres, the male age effect, and the younger brother effect. a, Relative *Wolbachia* titres in WT lines do not decrease with age. DNA copy number of the *w*Mel *groEL* gene is shown normalized to *D. melanogaster rp49* gene copy number in testes at the indicated ages. **b**, Absolute *Wolbachia* titres do not decrease from day 1 to day 7 males. **c, d**, In *w*Mel-infected males, WD0631 gene expression is equal between older (first day of emergence) and younger (fifth day of emergence) brothers while WD0632 gene expression is slightly higher in

early emerging brothers. **e**, There is no statistical difference in CI penetrance between older and younger brothers. $n = 8$ for each group in **a–d**; $n = 19–25$ for each group in **e**. Bars, mean ± s.d. *$P < 0.05$, ***$P < 0.001$, ****$P < 0.0001$ by ANOVA with Kruskal–Wallis test and Dunn's multiple test correction for **a**, **b**, and **e**, and two-tailed Mann–Whitney *U*-test for **c** and **d**. Exact *P* values are provided in Supplementary Table 7. These experiments were performed once.

a
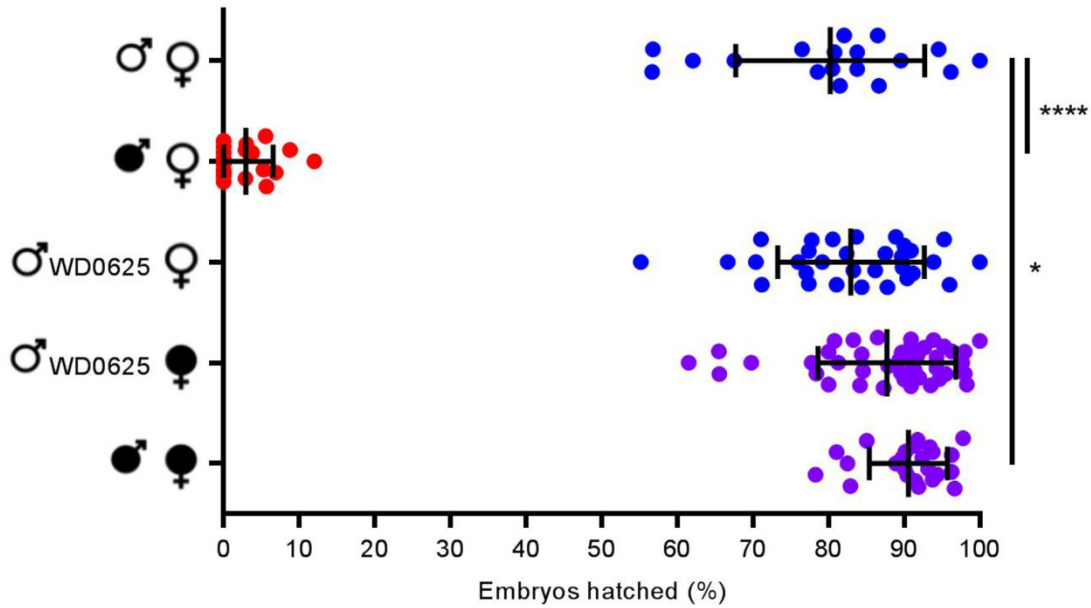
b
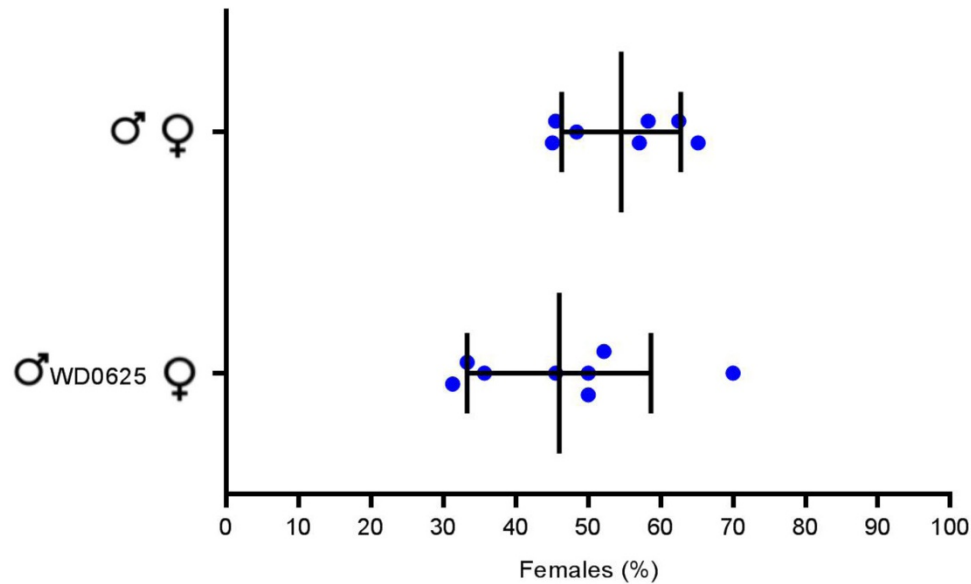
Extended Data Figure 5 | WD0625 transgene expression does not induce CI-like defects. Expression of control gene WD0625 in 1-day-old uninfected males does not affect (**a**) embryo hatch rates or (**b**) sex ratios. Infection status is designated with filled symbols for a *w*Mel-infected parent or open symbols for an uninfected parent. Transgenic flies are labelled with their transgene to the right of their male/female symbol. Unlabelled symbols represent WT flies. Data points are coloured according to the type of cross: blue, no CI; red, a CI cross; purple, a rescue cross with *w*Mel-infected females. $n = 18–47$ for each group in **a**; $n = 7–8$ for **b**. Bars, mean ± s.d. *$P < 0.05$, ***$P < 0.001$ by ANOVA with Kruskal–Wallis test and Dunn's multiple test correction. Exact P values are provided in Supplementary Table 7. This experiment was replicated three times.
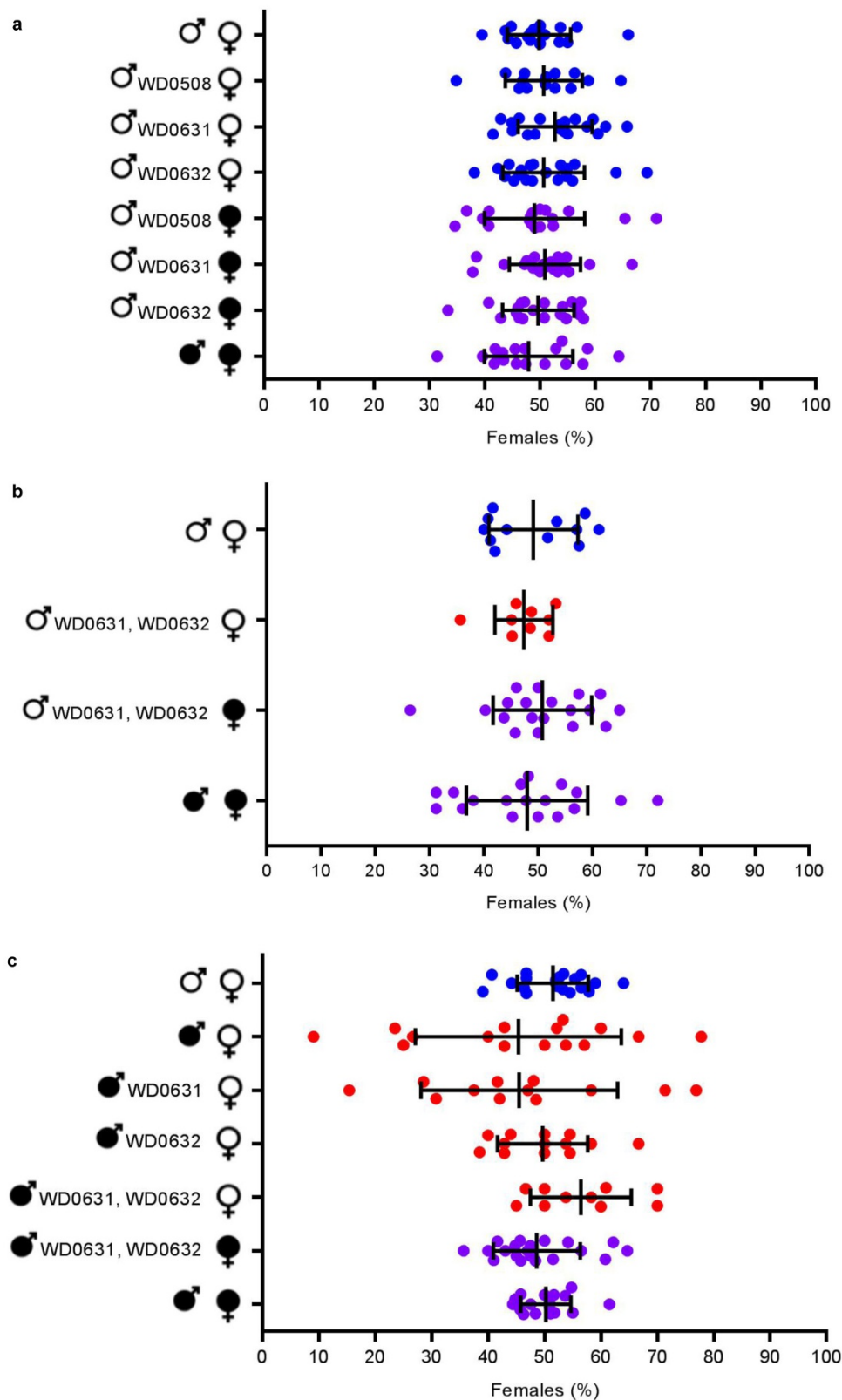
Extended Data Figure 6 | Expression of transgenes does not alter sex ratios. Graphs correspond to the same crosses as in Fig. 3. Infection status is designated with filled symbols for a *w*Mel-infected parent or open symbols for an uninfected parent. Transgenic flies are labelled with their transgene to the right of their gender symbol. Unlabelled gender symbols represent WT flies. Data points are coloured according to the type of cross: blue, no CI; red, a CI cross; purple, a rescue cross with *w*Mel-infected females. $n = 10$–36 for each group. Bars, mean ± s.d. Statistics include a Kruskal–Wallis tests and Dunn's multiple test corrections. The experiment in Extended Data Fig. 6a, c was performed once, while that in Extended Data Fig. 6b was performed twice.
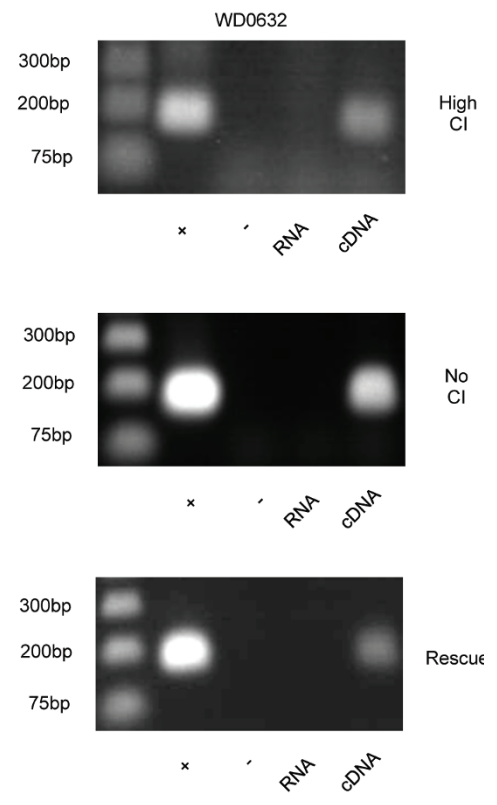
a

WD0508

b

WD0625

c

WD0631

d

WD0632
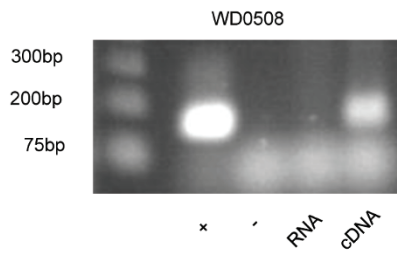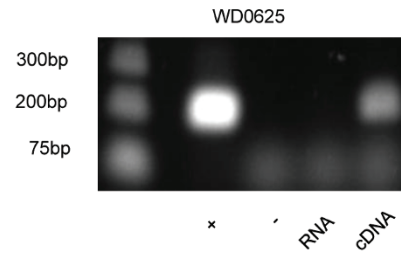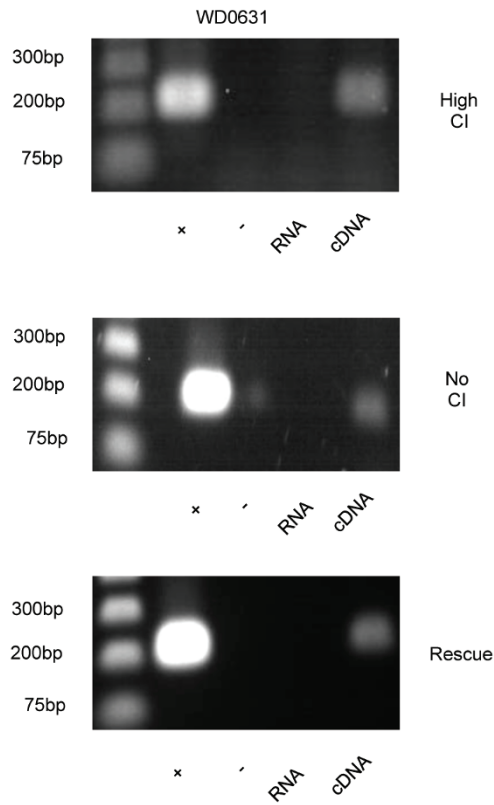
Extended Data Figure 7 | Transgenes are expressed in testes.
a, **b**, WD0508 and WD0625 transgenes are expressed in testes as evident by PCR performed against cDNA generated from dissected males used in Fig. 3a and Extended Data Fig. 5a, respectively. **c**, **d**, WD0631 and WD0632 transgenes are expressed in the testes from transgenic males specifically inducing high CI, no CI, or rescued CI. Testes were removed from males used in a replicate of Fig. 3b. $n = $ six pools of six pairs of testes, with representative image shown. This experiment was performed once.
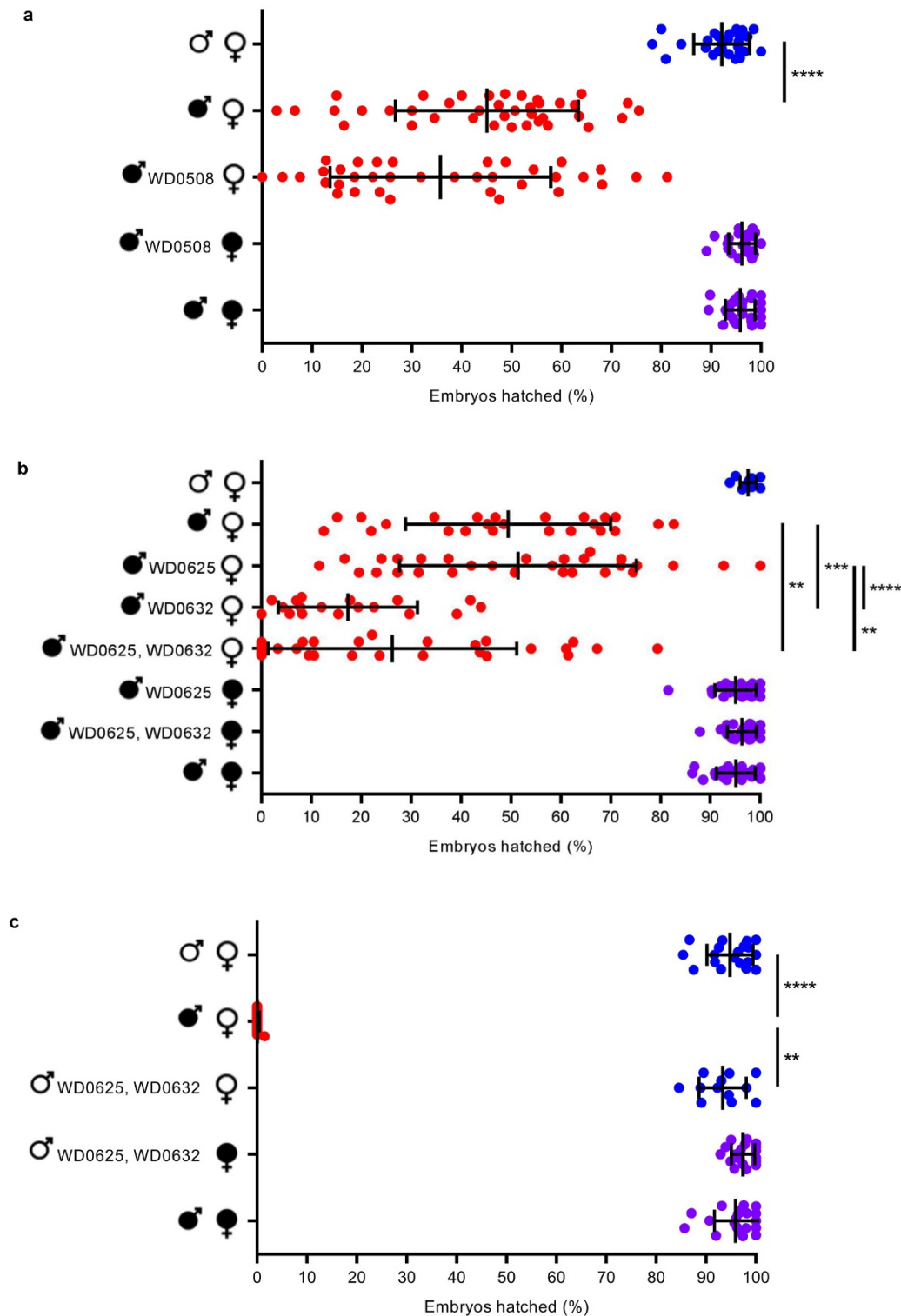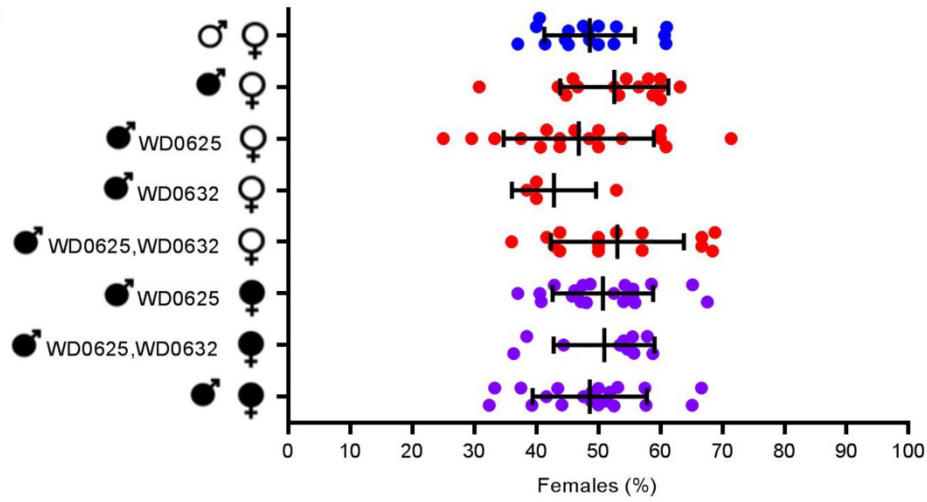
Extended Data Figure 8 | Transgenic expression of WD0508, WD0625, and WD0625/WD0632 (*cifB*) does not enhance or induce CI. a, The WD0508 transgene alone does not enhance CI in 2- to 4-day-old infected males. **b**, The WD0625 transgene alone does not enhance CI either; conversely, WD0632 does enhance CI as previously shown in Fig. 3c. The WD0625 transgene together with WD0632 does not enhance CI further than WD0632 alone. **c**, WD0625/WD0632 dual expression cannot induce CI in uninfected 1-day-old males. Infection status is designated with filled symbols for a *w*Mel-infected parent or open symbols for an uninfected parent. Transgenic flies are labelled with their transgene to the right of their male/female symbol. Unlabelled symbols represent WT flies. Data points are coloured according to the type of cross: blue, no CI; red, a CI cross; purple, a rescue cross with *w*Mel-infected females. *n* = 12–44 for each group. Bars, mean ± s.d. **P < 0.01, ***P < 0.001, ****P < 0.0001 by ANOVA with Kruskal–Wallis test and Dunn's multiple test correction. Exact *P* values are provided in Supplementary Table 7. These experiments were done twice (**a**, **c**), three times (**b**, WD0625, WD0632), or once (**b**, WD0625/WD0632).
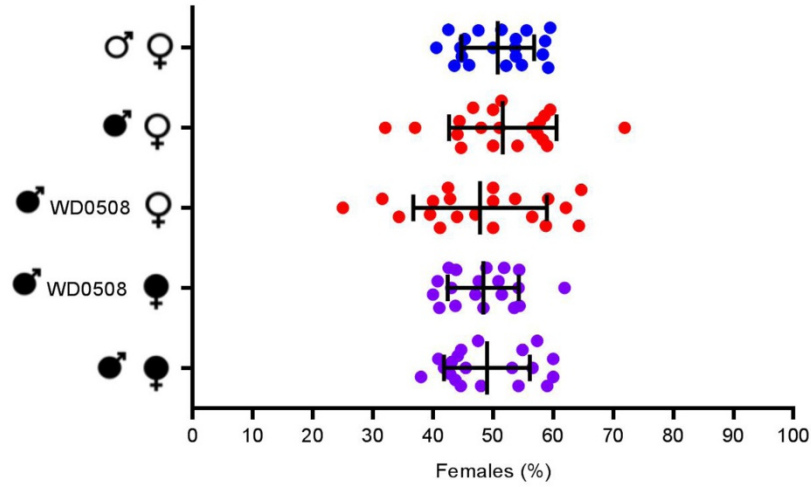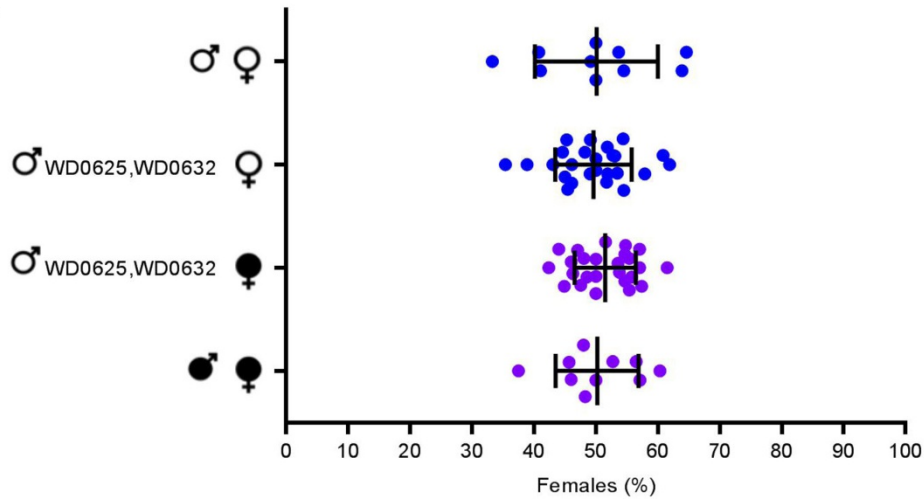
Extended Data Figure 9 | Transgenic expression of control genes does not affect sex ratios. All flies are from same crosses shown in Extended Data Fig. 8, except for **c**, which comes from a replicate experiment. Infection status is designated with filled symbols for a *w*Mel-infected parent or open symbols for an uninfected parent. Transgenic flies are labelled with their transgene to the right of their male/female symbol.

Unlabelled symbols represent WT flies. Data points are coloured according to the type of cross: blue, no CI; red, a CI cross; purple, a rescue cross with *w*Mel-infected females. $n = 4$–27 for each group. Bars, mean $\pm$ s.d. Statistics performed by ANOVA with Kruskal–Wallis test and Dunn's multiple test correction. These experiments were done twice (**b**) or once (**a**, **c**).
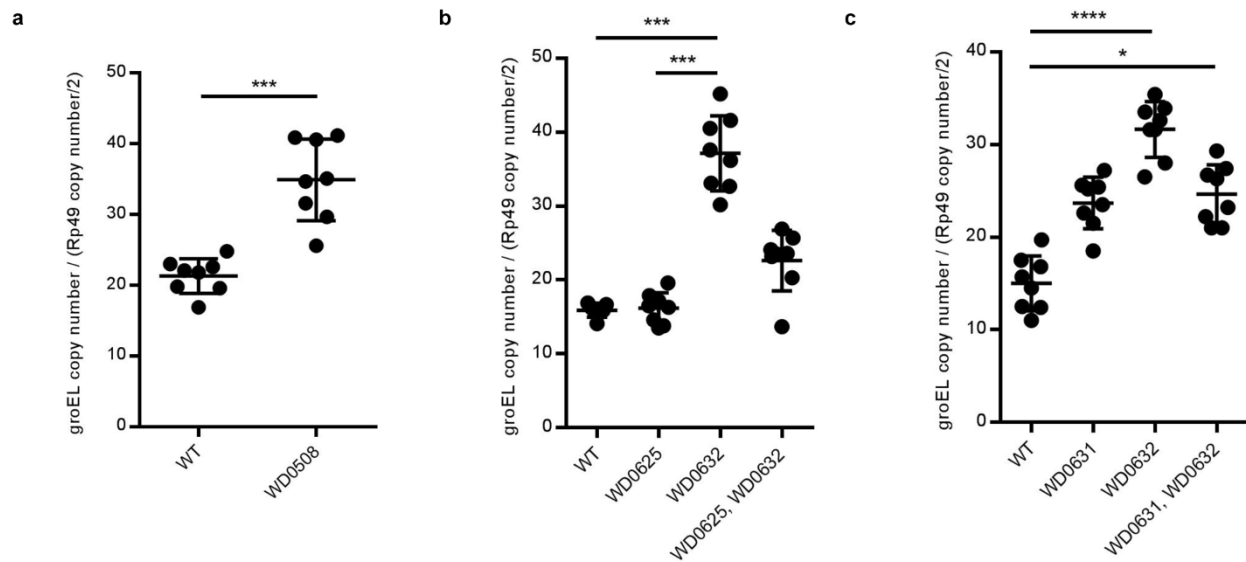
Extended Data Figure 10 | There is variation in *Wolbachia* titres in transgenic lines. a–c, Relative *Wolbachia* titres are higher in WD0508, WD0631, and WD0632 (*cifB*) transgenic lines than in WT lines. This does not occur in the WD0625 transgenic line, nor does there appear to be an additive effect. DNA copy number of the *w*Mel *groEL* gene is shown normalized to *D. melanogaster rp49* gene copy number in testes of the indicated strains. $n = 8$ independent pools of 15 pairs of testes for each group. Bars, mean ± s.d. *$P < 0.05$, ***$P < 0.001$, ****$P < 0.0001$ for two-tailed Mann–Whitney U-test (**a**) and Kruskal–Wallis test with Dunn's multiple test correction (**b**, **c**). Exact P values are provided in Supplementary Table 7. These experiments were done once.