

Calibrating and documenting host-switching and evolution of incompatibility loci for two closely related *Wolbachia* clades

J. Dylan Shropshire ^{1,2,*} William R. Conner ¹ Dan Vanderpool ³ Ary A. Hoffmann ⁴ Michael Turelli ^{5,*}
Brandon S. Cooper ^{1,*}

¹Division of Biological Sciences, University of Montana, Missoula, MT 59812, United States

²Department of Biological Sciences, Lehigh University, Bethlehem, PA 18015, United States

³Forest Service, National Genomics Center for Wildlife and Fish Conservation, Missoula, MT 59801, United States

⁴Pest and Environmental Adaptation Research Group, School of BioSciences, Bio21 Institute, University of Melbourne, Parkville, Victoria 3010, Australia

⁵Department of Evolution and Ecology, University of California, Davis, CA 95616, United States

*Corresponding authors: J. Dylan Shropshire. Email: shropshirejd@lehigh.edu; Michael Turelli. Email: mturelli@ucdavis.edu; Brandon S. Cooper. Email: brandon.cooper@umontana.edu

Maternally inherited *Wolbachia* alphaproteobacteria are the most common arthropod endosymbionts. Often, *Wolbachia* spread to high frequencies through cytoplasmic incompatibility, in which *cif* loci act through sperm to kill embryos lacking *Wolbachia*. Closely related *Wolbachia* with diverse *cif* loci often associate with anciently diverged hosts, but the timescale of associations remains uncertain. We produce new calibrations based on filarial nematodes with vertically inherited *Wolbachia* that codiverge with their hosts. Applying these calibrations to *Wolbachia* variants closely related to pathogen-blocking wMel from *Drosophila melanogaster*, we demonstrate that over a timescale of 1 to 2 million years, a core set of single-copy *Wolbachia* loci evolve largely through bifurcation rather than by gene exchange with distant *Wolbachia*. Dating bifurcating core genomes, we show that “wMel-like” *Wolbachia* diverged 2.1×10^5 to 2.4×10^6 years inhabit dipteran and hymenopteran hosts diverged more than 10^8 years. Previous published analysis of variants related to wRi from *Drosophila simulans*, the first *Wolbachia* found in a drosophilid, concluded that “wRi-like” *Wolbachia* spread among different *Drosophila* in tens of thousands of years. However, our new calibrations suggest these estimates from a mutation-based calibration underestimated wRi-like spread by about a factor of 7. In addition, *cif* exchanges between wMel-like and wRi-like *Wolbachia* genomes have occurred over $\sim 10^4$ to 10^6 years. Comparing intact *cif* loci found in various *Wolbachia*, we find function-preserving selection in their evolution. We discuss these results in light of theoretical predictions concerning selection on cytoplasmic incompatibility phenotypes within and among host lineages. The wMel variants analyzed may offer new options for *Wolbachia*-based biocontrol efforts.

Keywords: cytoplasmic incompatibility; *Drosophila*; endosymbiosis; genome evolution; horizontal transmission; host switching; selfish elements

Introduction

Maternally inherited *Wolbachia* bacteria were first discovered in the ovaries of the mosquito *Culex pipiens* (Hertig and Wolbach 1924). *Wolbachia* are now estimated to occur in about half of all terrestrial arthropod species (Weinert et al. 2015), as well as some nematodes (Ferri et al. 2011). Some clades of filarial nematodes require *Wolbachia* for normal development and fertility (Bandi et al. 1998; Manoj et al. 2021); in contrast, variants associated with arthropods tend to be facultative and not required by hosts (Hoffmann and Cooper 2025). Most facultative *Wolbachia* show phylogenetic distributions consistent with horizontal host switching, with closely related variants observed in distantly related hosts (O'Neill et al. 1992; Turelli et al. 2018; Vancaester and Blaxter 2023). Using clades defined by their affinity to the first two *Wolbachia* variants found in *Drosophila* species—wMel in *Drosophila melanogaster* (Hoffmann 1988) and wRi in *Drosophila*

simulans (Hoffmann et al. 1986)—our goal is to describe the tempo and patterns of their distribution among diverse hosts and their genome evolution, focusing on pairs of “cytoplasmic incompatibility factor” (*cif*) genes.

Both wRi and wMel (and many *Wolbachia* in our “wMel-like” and “wRi-like” samples) cause cytoplasmic incompatibility (CI), the most common reproductive phenotype associated with *Wolbachia* (Shropshire et al. 2020a). Embryos from females lacking *Wolbachia* fertilized by sperm from males with *Wolbachia* suffer increased mortality, the defining feature of CI. The genes that cause CI (*cifB* and occasionally *cifA*) and rescue CI-induced lethality (*cifA*) are associated with prophage (Wovirus) regions of *Wolbachia* genomes (Beckmann et al. 2017; LePage et al. 2017; Shropshire et al. 2018, 2021a; Shropshire and Bordenstein 2019; Adams et al. 2021; Sun et al. 2022). Woviruses are diverse and show patterns consistent with horizontal movements among *Wolbachia* genomes (Masui et al. 2000; Bordenstein and Wernegreen 2004; Gavotte

et al. 2007; Chafee et al. 2010; Bordenstein and Bordenstein 2022), analogous to *Wolbachia* host switching (O'Neill et al. 1992; Boyle et al. 1993; Werren et al. 1995; Heath et al. 1999; Van Meer and Stouthamer 1999; Dyson et al. 2002; Russell et al. 2009; Ahmed et al. 2016; Turelli et al. 2018; Martinez et al. 2021; Vancaester and Blaxter 2023). *cif* diversity also reflects horizontal movements (Lindsey et al. 2018; Shropshire et al. 2020a; Martinez et al. 2021; Vancaester and Blaxter 2023), including patterns consistent with *cif* movements between *Woviruses* (LePage et al. 2017; Cooper et al. 2019). Transposable elements are associated with *cif* genes within *Wovirus* regions (Amoros et al. 2025a), plausibly mediating these exchanges (Gillespie et al. 2018; Cooper et al. 2019; Madhav et al. 2020). Hence, *Wolbachia* move among hosts, *Woviruses* move among *Wolbachia* genomes, and *cifs* move with and among *Woviruses*. The timescales of these movements are uncertain.

Understanding the timescale of *Wolbachia* acquisitions and genomic evolution requires accurate calibrations of *Wolbachia* divergence times. Turelli et al. (2018) estimated divergence times for *wRi*-like *Wolbachia* using a mutation-based calibration of *Wolbachia* molecular evolution from Richardson et al. (2012). They concluded that variants that diverged only about 30,000 years ago occupy *Drosophila* hosts that diverged over 40 million years ago. However, mutation rates may overestimate the substitution rates that underlie chronograms if mutation rates do not accurately reflect substitution rates (Richardson et al. 2012). Fossil-based estimates of divergence times provide an alternative framework for estimating rates of genome evolution. When *Wolbachia* codiverge with their hosts, fossil-based host divergence times match divergence times for their concordantly diverging *Wolbachia*, allowing estimation of *Wolbachia* molecular evolution rates (cf. Raychoudhury et al. 2009; Gerth and Bleidorn 2016). However, *Wolbachia*-host codivergence is rarely documented by comparing molecular divergence of nuclear and mtDNA genomes of hosts with that of their *Wolbachia* (cf. Hamm et al. 2014; Conner et al. 2017). Hence, it is often difficult to establish *Wolbachia*-host codivergence for facultative variants with uncertain evolutionary histories. In addition, chronograms, which are essential to conclusions, assume bifurcating evolution; however, prior analyses suggest that *Wolbachia* can have a mosaic genome structure produced by recombination and horizontal genomic exchanges between closely and distantly related variants (Jiggins et al. 2001; Werren and Bartos 2001; Baldo et al. 2006; Ellegaard et al. 2013; Wang et al. 2020). Prior attempts to resolve the timescales of *Wolbachia* movements (Raychoudhury et al. 2009; Gerth and Bleidorn 2016; Turelli et al. 2018) have mostly ignored the potential for recombination to bias divergence-time estimates.

To produce robust calibrations, we first assess support for the bifurcating evolution of regions of *Wolbachia* genomes. We provide evidence that over the timescales of divergence of the *wMel*-like and *wRi*-like clades, large portions of their genomes, specifically most single-copy protein-coding genes, evolve as bifurcating lineages, with little evidence of horizontal acquisition of alleles from distantly related *Wolbachia*. We address how the recent bifurcation of these genomic regions can be reconciled with the evidence for recombination among *Wolbachia*. We use these bifurcating regions of genomes as a temporal standard to calibrate rates of *Wolbachia* molecular evolution using new data from trios of filarial nematodes (Qing et al. 2025) that carry obligate *Wolbachia* (Hoerauf et al. 2000; Fenn and Blaxter 2004; Taylor et al. 2005). Extensive evidence supports codivergence of filarial nematodes and their *Wolbachia* in particular clades (Moran 2007; Lefoulon et al. 2016; Vancaester et al. 2025). The

case for codivergence is strengthened by data supporting obligate symbioses (Manoj et al. 2021).

We provide new estimates of *wRi*-like divergence and also estimate divergence times for closely related *wMel*-like *Wolbachia* from widely divergent hosts, spanning holometabolous insects. Although the nematode *Wolbachia* lineages we use for calibration are distantly related to *wMel*-like and *wRi*-like *Wolbachia*, the molecular evolution rates they imply are consistent with some previous estimates derived from the facultative and plausibly codiverging *Wolbachia* in *Nasonia* wasps (Raychoudhury et al. 2009) and *Nomada* bees (Gerth and Bleidorn 2016) that are much more similar to *wMel* and *wRi* than the nematode *Wolbachia* lineages. With our calibrations, we also show that *Woviruses* and *cif* genes are rapidly gained and lost outside of the bifurcating *wRi*-like and *wMel*-like genomic regions. Analyzing selection on intact *cif* genes from *Wolbachia* from different host species, we find evidence that selection acts on both *cifA* and *cifB* sequence evolution. We discuss these results in light of faster *cifB* mutation accumulation (Meany et al. 2019; Martinez et al. 2021) and theoretical predictions concerning how selection acts on CI phenotypes within host species (Prout 1994; Turelli 1994; Haygood and Turelli 2009) and in the proliferation of CI phenotypes among *Wolbachia* across host species (Hurst and McVean 1996; Turelli et al. 2022). Overall, our analyses provide insights into *Wolbachia* genome evolution and shed light on the timescale of host switching and the evolution of incompatibility loci for two closely related model *Wolbachia* clades.

Results

wMel-like *Wolbachia* and their hosts

Our study includes 19 closely related *Wolbachia* plus the model pathogen-blocking *wMel* found in *D. melanogaster*. Figure 1a presents our cladogram for these variants, which we refer to as “*wMel*-like” *Wolbachia* (Hague et al. 2020). Hosts that carry *wMel*-like *Wolbachia* are diverse taxonomically and in their life histories (Table 1 & Supplementary Table 1; Supplementary Information). They include species from the clades Diptera and Hymenoptera, which diverged approximately 350 million years ago (MYA), and span the Holometabola (Wang et al. 2016; Fig. 1b; Supplementary Information). These patterns imply phylogenetic discordance between *wMel*-like *Wolbachia* and their hosts, as observed between many *Wolbachia* and host species (Fig. 2b in Vancaester and Blaxter 2023; O'Neill et al. 1992), including *wRi*-like *Wolbachia* found in hosts that span the *Drosophila* genus (Turelli et al. 2018).

After introducing host chronograms for *Wolbachia* calibrations and establishing that major portions of *Wolbachia* genomes show patterns consistent with a bifurcating evolutionary history (i.e. a history of vertical descent), we use new data on filarial nematodes with vertically inherited *Wolbachia* to resolve the timescale of *wMel*-like *Wolbachia* host switching and revise our prior *wRi*-like estimates. We then describe the tempo and patterns of *wMel*-like and *wRi*-like *Wolbachia* genome evolution outside of bifurcating regions, focusing on pairs of *cif* genes involved in protecting from and inducing CI. Our study does not include all known variants, for example, variants closely related to *wAu* (Miller and Riegler 2006), but our conclusions are robust to incomplete sampling of these clades as we discuss below.

Host chronograms for calibrating *Wolbachia* divergence

Wolbachia chronograms rely on molecular clock approximations that can be obtained when *Wolbachia* codiverge with their hosts,

Table 1. Our study includes 20 wMel-like and the 8 wRi-like *Wolbachia* ($N = 37$ wRi-like genomic variants) described in Turelli et al. (2018).

<i>Wolbachia</i>	Group	Host Order	Host Genus	Host Species
wSan	wMel-like	Diptera	<i>Drosophila</i>	<i>santomea</i>
wYak	wMel-like	Diptera	<i>Drosophila</i>	<i>yakuba</i>
wTei	wMel-like	Diptera	<i>Drosophila</i>	<i>teissieri</i>
wZta	wMel-like	Diptera	<i>Zapionus</i>	<i>taronus</i>
wZts	wMel-like	Diptera	<i>Zapionus</i>	<i>tsacasi</i>
wMel	wMel-like	Diptera	<i>Drosophila</i>	<i>melanogaster</i>
wAu	wMel-like	Diptera	<i>Drosophila</i>	<i>simulans</i>
wTro	wMel-like	Diptera	<i>Drosophila</i>	<i>tropicalis</i>
wBor	wMel-like	Diptera	<i>Drosophila</i>	<i>borealis</i>
wInc	wMel-like	Diptera	<i>Drosophila</i>	<i>incompta</i>
wAch	wMel-like	Diptera	<i>Drosophila</i>	<i>sp. aff. chauvacae</i>
wBocq	wMel-like	Diptera	<i>Drosophila</i>	<i>bocqueti</i>
wTris	wMel-like	Diptera	<i>Drosophila</i>	<i>tristis</i>
wRec	wMel-like	Diptera	<i>Drosophila</i>	<i>recens</i>
wMal	wMel-like	Diptera	<i>Drosophila</i>	<i>malagassya</i>
wSeg	wMel-like	Diptera	<i>Drosophila</i>	<i>seguyi</i>
wSpa	wMel-like	Diptera	<i>Scaptomyza</i>	<i>pallida</i>
wAra	wMel-like	Diptera	<i>Drosophila</i>	<i>arawakana</i>
wSbr	wMel-like	Diptera	<i>Sphyracephala</i>	<i>brevicornis</i>
wDal	wMel-like	Hymenoptera	<i>Diachasma</i>	<i>alloeum</i>
wRi	wRi-like	Diptera	<i>Drosophila</i>	<i>simulans</i>
wAno	wRi-like	Diptera	<i>Drosophila</i>	<i>anomala</i>
wPan	wRi-like	Diptera	<i>Drosophila</i>	<i>pandora</i>
wAna	wRi-like	Diptera	<i>Drosophila</i>	<i>ananassae</i>
wSuz	wRi-like	Diptera	<i>Drosophila</i>	<i>suzukii</i>
wSpc	wRi-like	Diptera	<i>Drosophila</i>	<i>subpulchrella</i>
wAur	wRi-like	Diptera	<i>Drosophila</i>	<i>auraria</i>
wTri	wRi-like	Diptera	<i>Drosophila</i>	<i>triauraria</i>
wNfla	Nomada	Hymenoptera	<i>Nomada</i>	<i>flava</i>
wNfe	Nomada	Hymenoptera	<i>Nomada</i>	<i>ferruginata</i>
wNleu	Nomada	Hymenoptera	<i>Nomada</i>	<i>leucophthalma</i>
wNpa	Nomada	Hymenoptera	<i>Nomada</i>	<i>panzeri</i>
wOo	supergroup C	Rhabditida	<i>Onchocerca</i>	<i>ochengi</i>
wOv	supergroup C	Rhabditida	<i>Onchocerca</i>	<i>volvulus</i>
wDimm	supergroup C	Rhabditida	<i>Dirofilaria</i>	<i>immitis</i>
wBm	supergroup D	Rhabditida	<i>Brugia</i>	<i>malayi</i>
wBp	supergroup D	Rhabditida	<i>Brugia</i>	<i>pahangi</i>
wWb	supergroup D	Rhabditida	<i>Wuchereria</i>	<i>bancrofti</i>
wLs	supergroup D	Rhabditida	<i>Litomosoides</i>	<i>sigmodontis</i>

All wMel-like and wRi-like variants are members of the larger *Wolbachia* supergroup A. We also include the supergroup A *Nomada* ($N = 4$) and the supergroups C ($N = 3$) and D ($N = 4$) nematode-associated *Wolbachia* variants used for alternative calibrations. We present taxonomic information for both *Wolbachia* and the hosts that carry them. In [Supplementary Table 1](#), we provide all genome accessions and metrics of *Wolbachia* genome quality.

and fossil-based estimates of host divergence times are available. We first consider host chronograms for 3 cases in which *Wolbachia*-host codivergence has been proposed: the 4-species *Nomada* bee clade with facultative *Wolbachia*, analyzed by Gerth and Bleidorn (2016), and 2 trios of filarial nematodes with obligate *Wolbachia*, for which both *Wolbachia* and host genomic data have recently become available.

Figure 2a presents our relative chronogram for the 4-species *Nomada* clade: (*Nomada ferruginata*, *Nomada panzeri*, *Nomada flava*, and *Nomada leucophthalma*). Following Gerth and Bleidorn (2016, p. 4), the crown age of the clade is approximately 2.14 MYA. The internal nodes have relative ages 0.58 (with 95% HPD support interval 0.33 to 0.90) and 0.36 (with 95% HPD support interval 0.16 to 0.61). Thus, the MRCA ages (and support intervals) for (*N. panzeri*, *N. flava*, and *N. leucophthalma*) and (*N. flava*, *N. leucophthalma*) are roughly 1.24 (0.71 to 1.93) MYA and 0.77 (0.34 to 1.31) MYA. These estimates of internal node ages are consistent with those that Gerth et al. (2013, Fig. 3) obtained from more limited host nuclear data. In contrast, when Gerth and Bleidorn (2016, Fig. 3) estimated these node ages using *Wolbachia* genomes, assuming codivergence with all 4 hosts, their point estimates for these branch splits, 0.43 MYA and 0.25 MYA, respectively, fall outside the support intervals obtained from our more extensive nuclear-

gene analyses. We argue that these comparative genomic data, including *Wolbachia* and host nuclear and mtDNA, suggest horizontal transmission of *Wolbachia* within the ingroup clade (*N. panzeri*, (*N. flava* and *N. leucophthalma*)) and explain the discordant chronograms based on host vs *Wolbachia* genomes ([Supplementary Information](#)).

We also considered 2 trios of filarial nematodes, each with a pair of congeners and an outgroup containing a *Wolbachia* variant belonging to the same supergroup as those in the congeners: (*Dirofilaria immitis*, *Onchocerca volvulus*, and *Onchocerca ochengi*) that contain supergroup C *Wolbachia*, and (*Wuchereria bancrofti*, *Brugia malayi*, and *Brugia pahangi*) that contain supergroup D *Wolbachia*. Unlike prior calibrations from *Nasonia* and *Nomada* that carry facultative *Wolbachia*, these hosts carry obligate and vertically inherited *Wolbachia* known to codiverge with their hosts. For both trios, Qing et al. (2025) estimated the time to the MRCA of the 3 species. The estimate for *Dirofilaria* and *Onchocerca* is 64.73 MYA with 95% HPD 37.56 to 91.96 MYA; the estimate for *Wuchereria* and *Brugia* is 28.97 MYA, with 95% HPD 12.8 to 48.59 MYA. Figures 2b and 2c provide estimates for the divergence times of the congeners relative to their respective outgroup. The relative age of the MRCA for *O. volvulus* and *O. ochengi* is 0.031, with 95% HPD 0.012 to 0.059, corresponding to about 2.01 MYA;

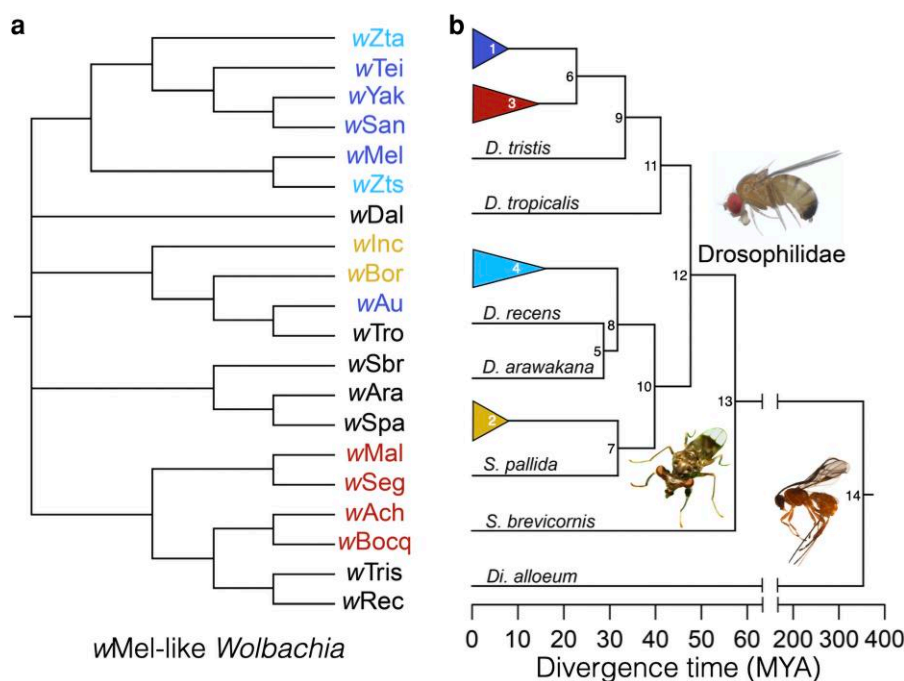


Fig. 1. Closely related wMel-like *Wolbachia* occupy divergent hosts. a) A cladogram for the wMel-like *Wolbachia* included in our study (Table 1 & Supplementary Table 1). Note that several of the basal nodes are not resolved, reflecting rapid diversification of these wMel-like variants. The *Wolbachia* labels are colored to match their hosts in panel b. b) An approximate chronogram for the hosts of the wMel-like *Wolbachia* from panel a. Node numbers correspond to the node numbers in Table 2, which provides node ages and approximate credible intervals. For the drosophilids, the dark blue clade includes *melanogaster* subgroup hosts, the red clade includes *montium* subgroup hosts, the light blue clade includes 2 *Zaprionus* hosts, and the gold clade includes *D. borealis* and *D. incompta*. The species within focal clades and their relationships are also presented in Table 2, and additional information on host divergence is presented in the Supplementary Information. Open-source images taken by CBD (*Diachasma alloeum*), K. Schulz (*S. brevicornis*) and T. Wheeler (*D. teissieri*).

Table 2. Ages and approximate confidence intervals in millions of years for hosts that carry wMel-like *Wolbachia*.

Node	Species and Clade Relationships	Point	95% CI
1	<i>melanogaster</i> subgroup: ((<i>D. melanogaster</i> , <i>D. simulans</i>), (<i>D. teissieri</i> , (<i>D. santomea</i> , <i>D. yakuba</i>)))	6.7	(5.8, 7.7)
2 ^a	(<i>D. borealis</i> , <i>D. incompta</i>)	8.2	(3.7, 14.5)
3 ^b	<i>montium</i> subgroup: (<i>D. malagassya</i> , <i>D. seguyi</i>), (<i>D. sp. aff. chauvacae</i> , <i>D. bocqueti</i>)	12.7	(11.4, 14.3)
4	<i>Zaprionus</i> clade: (<i>Z. taronus</i> , <i>Z. tsacasi</i>)	14.2	(12.3, 16.7)
5	(<i>D. recens</i> , <i>D. arawakana</i>)	20.4	(18.2, 22.7)
6	(<i>melanogaster</i> subgroup, <i>montium</i> subgroup)	22.2	(19.9, 24.6)
7	(<i>S. pallida</i> , (<i>D. borealis</i> , <i>D. incompta</i>))	30.1	(27.7, 32.7)
8	(<i>Zaprionus</i> clade, (<i>D. recens</i> , <i>D. arawakana</i>))	30.7	(28.9, 33.2)
9	(<i>D. tristis</i> , (<i>melanogaster</i> subgroup, <i>montium</i> subgroup))	32.8	(29.8, 35.8)
10	((<i>Zaprionus</i> clade, (<i>D. recens</i> , <i>D. arawakana</i>)), (<i>S. pallida</i> , (<i>D. borealis</i> , <i>D. incompta</i>)))	34.1	(31.6, 36.8)
11	(<i>D. tropicalis</i> , (<i>D. tristis</i> , (<i>melanogaster</i> subgroup, <i>montium</i> subgroup)))	40.3	(37.4, 43.2)
12	(<i>Sophophora</i> subgenera, <i>Drosophila</i> subgenera)	46.8	(43.9, 49.9)
13	(<i>Drosophilidae</i> , <i>Diopsoidea</i>)	59	(47, 70)
14	(<i>Diptera</i> , <i>Hymenoptera</i>)	350	(329, 378)

Node numbers correspond to those presented in Fig. 1b and are presented in order from the youngest to the oldest nodes. Supplementary Table 1 provides the genome accessions.

^aTo estimate the divergence time between *D. borealis* and *D. incompta*, we estimated a relative chronogram for (*D. hydei*, *D. virilis*), (*D. borealis*, *D. incompta*). *D. hydei* is the outgroup and an estimate for the crown age is available from Suvorov et al. (2022).

^bTo estimate the divergence time of *D. malagassya* and *D. seguyi*, and also *D. sp. aff. chauvacae* and *D. bocqueti*, we estimated a relative chronogram for (*D. jambulina*, (*D. bocqueti*, *D. sp. aff. chauvacae*), (*D. malagassya*, *D. seguyi*)). *D. jambulina* is the outgroup and an estimate for the crown age is available from Suvorov et al. (2022).

for *B. malayi* and *B. pahangi*, the relative divergence time is 0.23, with 95% HPD 0.09 to 0.41, corresponding to about 6.66 MYA.

Assessing support for bifurcating evolution of major portions of *Wolbachia* genomes across 2 orders of magnitude of genomic divergence

Robust calibrations rely on bifurcating phylogenies, which assume that differences among tip sequences arose from independent evolutionary changes along the branches, without introgressions from genomes with distinct evolutionary histories. Recombination

between *Wolbachia* genomes has been repeatedly demonstrated (Jiggins et al. 2001; Werren and Bartos 2001; Baldo et al. 2006; Ellegaard et al. 2013); hence, it is important to assess evidence for recombination and the extent to which horizontal acquisition of sequence elements may preclude meaningful analyses of divergence times. Prior attempts to resolve the timescales of *Wolbachia* movements (including our own) have mostly ignored the potential for recombination to bias divergence-time estimates.

To understand the phylogenetic implications of recombination, it is critical to recognize that recombination events in

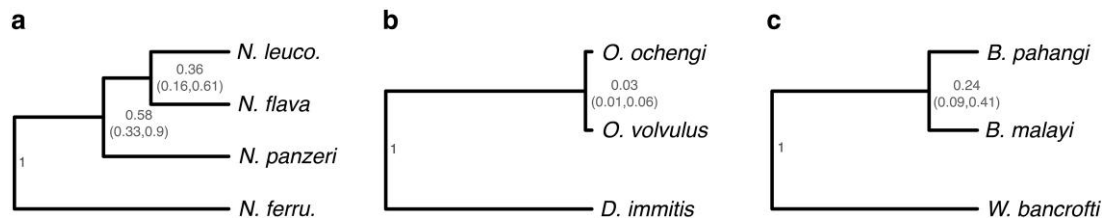


Fig. 2. Relative chronogram for the 4-species *Nomada* clade and estimates of divergence times of nematode congeners relative to their respective outgroup. a) A relative chronogram for the 4-species *Nomada* clade: (*N. ferruginata*, (*N. panzeri*, (*N. flava*, *N. leucophthalma*))). Following Gerth and Bleidorn (2016, p. 4), the crown age of the clade is approximately 2.14 MYA. The internal nodes have relative ages 0.58 (with 95% HPD support interval 0.33 to 0.90) and 0.36 (with 95% HPD support interval 0.16 to 0.61). Thus, the MRCA ages (and support intervals) for (*N. panzeri*, (*N. flava*, *N. leucophthalma*)) and (*N. flava*, *N. leucophthalma*) are roughly 1.24 (0.71 to 1.93) MYA and 0.77 (0.34 to 1.31) MYA. Estimates for the divergence times of b) *O. volvulus* and *O. ochengi* and of c) *B. malayi* and *B. pahangi*, each relative to their respective outgroup. The relative age of the MRCA for *O. volvulus* and *O. ochengi* is 0.031, with 95% HPD 0.012 to 0.059, corresponding to about 2.01 MYA; for *B. malayi* and *B. pahangi*, the relative divergence time is 0.23, with 95% HPD 0.09 to 0.41, corresponding to about 6.66 MYA.

bacteria are asymmetric and heterogeneous. For *Wolbachia*, “recombination” can result in (i) replacing a genome fragment in a recipient genome with a homologous fragment from a donor genome, (ii) integrating donor genome elements, such as transposable elements (TEs) or prophages, into recipient positions that need not be homologous to positions in the donor genome, or (iii) recombination-mediated rearrangements within a single genome, such as inversions, phage or TE movements, or duplications. The recombination studies cited above found portions of individual loci with distinct phylogenies, indicating horizontal acquisition of fragments from relatively distantly related donor genomes. The resulting mosaic structure is clearly inconsistent with a single bifurcating history of the recombinant locus. Similarly, horizontal acquisition of larger fragments, involving complete genes with alleles that are distantly related to the alleles replaced, is inconsistent with bifurcation. In contrast, exchange of genome fragments between essentially identical *wMel*-like or *wRi*-like genomes that involve few or no SNPs would have little or no effect on estimating an essentially bifurcating history. Because it is well known that TEs and prophage move within and among *Wolbachia* genomes (Masui et al. 2000; Bordenstein and Wernegreen 2004; Riegler et al. 2005; Cerveau et al. 2015) and that rearrangements, such as inversions, are common (Wu et al. 2004; Klasson et al. 2009; Cerveau et al. 2011; Ellegaard et al. 2013; Baiao et al. 2021), we focus on determining whether closely related genomes, such as our *wRi*-like and *wMel*-like samples, exhibit a bifurcating history for single-copy, protein-coding *Wolbachia* loci (i.e. loci not within TEs or prophages). Intragenomic rearrangements, such as inversions, do not alter the bifurcating history of individual genes.

To better understand the dynamics of *Wolbachia* genome evolution in our system, we assessed evidence for bifurcating evolution in sets of variants with increasing levels of sequence divergence: (i) a set of 8 *wRi*-like *Wolbachia* (Turelli et al. 2018), (ii) the set of 20 *wMel*-like *Wolbachia* introduced here, and (iii) a broader set of 7 supergroup A variants. These 7 include 2 *wMel*-like variants (*wMel* and *wAu*), 2 *wRi*-like variants (*wRi* and *wTri*), *wHa* from *D. simulans*, *wBic* from *D. bicornuta*, and *wBar* from *D. barbarae*. As noted in Fig. 3, the clade defined by these 7 spans our *wRi*-like and *wMel*-like clades. In every case, our analyses included single-copy genes extracted from each set, using criteria described in our Materials and Methods. We focused on these genes because they are used in our downstream calibrations. In total, we extracted 525 genes (506,307 bp) from *wRi*-like genomes, 438 genes (409,848 bp) from *wMel*-like genomes, and 442 genes (413,115 bp) from the 7 supergroup A variants. We also assessed

whether the nematode *Wolbachia* genomes evolved primarily through bifurcation. In the following subsections, we describe the results of several tests demonstrating that these single-copy genes comprise bifurcating genomic regions, from which we can robustly estimate *Wolbachia* divergence times.

Analyzing intragenic recombination. We first assessed evidence for intralocus recombination using a genetic algorithm for recombination detection (GARD) (Kosakovsky Pond et al. 2006), plus all 3 statistical methods implemented in PhiPack (Bruen et al. 2006). We found no evidence for intragenic recombination in our least-diverged set of *wRi*-like *Wolbachia*. It is worth noting that given the extreme similarity of the *wRi*-like genomes, recombination between variants within this clade would be essentially impossible to detect, and our ability to detect recombination increases with the divergence of our sets of variants. However, we found almost no evidence for intragenic recombination in our set of slightly more diverged *wMel*-like *Wolbachia* (GARD: 0/411 genes; PHI test: 0/411, Max χ^2 test: 1/411, Neighbor Similarity Score: 0/411, FDR-adjusted $P < 0.01$), and in our most diverged set that includes 4 samples of supergroup D *Wolbachia* (GARD: 0/356 genes; PHI test: 0/356, Max χ^2 test: 1/356, Neighbor Similarity Score: 0/356, FDR-adjusted $P < 0.01$). Analysis of supergroup C *Wolbachia* was not possible due to an insufficient number of available sequences. In contrast, we found evidence of intragenic recombination within many of the genes that met our criteria for the intermediately diverged set of 7 supergroup A variants (GARD: 187/401 genes; PHI test: 40/401, Max χ^2 test: 110/401, Neighbor Similarity Score: 63/401, FDR-adjusted $P < 0.01$). In the Supplementary Information, we describe that when sufficient variation exists for GARD to produce a tree, its recombination calls are error prone if there is insufficient variation to accurately resolve trees for putative partitions. Thus, we completed several alternative tests that illustrate the bifurcating evolution of major portions of *Wolbachia* genomes across levels of divergence relevant to our analyses of *wMel*-like and *wRi*-like variants.

Analyzing single-gene phylogenies. To determine the prevalence of horizontal exchanges and departures from purely bifurcating evolution, we next assessed single-gene tree discordance. We estimated single-gene phylogenies from 442 individual genes in the group of 7 supergroup A *Wolbachia*, of which 52 genes produced fully resolved topologies with posterior probabilities > 0.95 at each node. Of the 52 phylogenetically informative loci, all supported the pairs (*wMel*, *wAu*) and (*wRi*, *wTri*) as closest relatives. Using IQ-TREE (Wong et al. 2025), we observed gene concordance factors (gCFs) > 0.99 for both (*wAu*, *wMel*) and (*wRi*, *wTri*) across bootstrap thresholds (50%, 60%, 80%, and 95%)

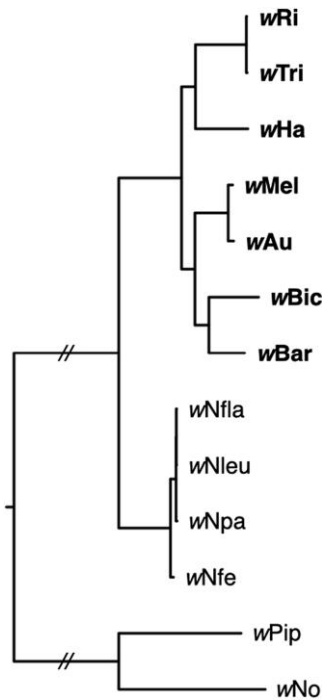


Fig. 3. Phylogram of 7 focal supergroup A *Wolbachia*, 4 supergroup A *Nomada*-derived *Wolbachia*, and *wNo* and *wPip* from supergroup B. Phylogram based on 177 concatenated loci (146,619 bp). The *Wolbachia* include 4 supergroup A variants from *Nomada* species, (*wNfla*, *wNleu*, *wNfe*, and *wNpa*), and *wNo* and *wPip* from supergroup B. These 6 *Wolbachia* were used to root the 7 focal supergroup A genomes (bold) that span the divergence of the *wMel*-like and *wRi*-like clades. The resulting “consensus phylogeny,” denoted C, for the 7 supergroup A variants is: (((*wMel*, *wAu*), (*wBic*, *wBar*)), ((*wRi*, *wTri*), *wHa*)). This phylogeny has posterior probabilities of 1 for all nodes and 100% bootstrap support over 1,000 replicates. Branches leading to supergroups A and B are shortened for presentation (//).

used to collapse gene-tree nodes prior to scoring decisiveness (Supplementary Table 2). These results suggest that essentially no horizontal transmission disrupts the bifurcating evolution of these genomic regions within the *wMel*-like and *wRi*-like clades. This is supported by our polymorphism analyses in the following subsection, which assessed the 52 loci described above and those in the unresolved class.

Departures from pervasive bifurcation seem more plausible when we consider all 7 supergroup A genomes that span the *wMel*-like and *wRi*-like clades. We estimated a consensus phylogeny by concatenating all 442 loci, using the 4 *Nomada*-derived *Wolbachia* as outgroups from supergroup A and *wNo* and *wPip* from supergroup B to root it. This “consensus phylogeny,” denoted C, is: (((*wMel*, *wAu*), (*wBic*, *wBar*)), ((*wRi*, *wTri*), *wHa*)), with posterior probabilities of 1 for all nodes and 100% bootstrap support over 1,000 replicates (Fig. 3). We also estimated an unrooted species tree from our set of 442 gene trees using weighted ASTRAL (Zhang and Mirarab 2022). The species tree agreed with C, with local posterior support of 1 for all internal branches. ASTRAL quartet frequencies were skewed toward the reference resolution (median $q_1 = 0.70$, range: 0.44 to 0.97; Supplementary Table 3), consistent with short but positive internode lengths under the multi-species coalescent. While gCFs were over 0.99 for (*wAu*, *wMel*) and (*wRi*, *wTri*), they were ~0.56 for the other nodes that were again relatively insensitive to the collapse threshold (Supplementary Table 2). It is notable that the stem branches

leading to the 3 clades (*wBic*, *wBar*), ((*wMel*, *wAu*), (*wBic*, *wBar*)), and ((*wRi*, *wTri*), *wHa*) are significantly shorter than the terminal branches leading to *wHa*, *wBic*, and *wBar* and the internal branches leading to the *wMel*-like and *wRi*-like pairs. Supplementary Table 4 indicates the number of loci that support each of the 15 possible unrooted trees of the 5 entities *wHa*, *wBic*, *wBar*, (*wMel*, *wAu*), and (*wRi*, *wTri*). Each topology was supported by at least 1 of the 52 phylogenetically informative loci, with 14 loci supporting the consensus phylogeny. Each of the alternative unrooted topologies, denoted A1 to A14 in Supplementary Table 4, is supported by 6 or fewer loci.

The loci supporting the alternative topologies, A1 to A14, are candidates for belonging to horizontally acquired genomic fragments. However, alternative topologies may also arise if multiple *Wolbachia* genomes that differ slightly are horizontally transmitted between host species. This could produce discordant gene trees by a process analogous to incomplete lineage sorting (Gillespie and Langley 1979; Maddison 1997; Edwards 2009). Indeed, the resolved ASTRAL species tree and the observed gCF values together point to short-branch ILS or inference errors associated with model mis-specification as plausible sources of discordance, rather than pervasive horizontal exchange. In contrast, if the “discordant” loci derive from recombination events involving integration of genomic fragments acquired via transformation or transduction from distantly related *Wolbachia* genomes, the loci might be expected to show atypically high levels of pairwise differences relative to loci that did not experience horizontal transfer (cf. Wang et al. 2020). In general, the phylogenetically informative loci might show relatively high levels of sequence divergence, given that we selected them based on their ability to produce a fully resolved tree. Moreover, if horizontal transfer involves large genomic segments, we might expect the discordant loci to be clustered. Supplementary Table 5 shows the positions of the 52 phylogenetically informative loci with respect to 4 reference genomes, *wMel*, *wAu*, *wRi*, and *wHa*. The genes are listed in their order of occurrence relative to the origin of the *wMel* reference genome. Under each alternative reference genome, we present in parentheses their order of appearance in that reference genome. The gene-order differences illustrate the pervasiveness of rearrangements, especially inversions, as documented by Ellegaard et al. (2013). The last 2 columns give 2 measures of gene-specific divergence among the 7 genomes: the fraction of polymorphic sites within each gene and the total number of substitutions across the entire tree (estimated from the single-gene phylogram).

As expected, the loci that supported non-consensus topologies vs the consensus topology have more polymorphic sites (means $\pm$ standard deviations: 0.051 ± 0.020 vs 0.040 ± 0.026) and higher estimated numbers of substitutions (0.062 ± 0.026 vs 0.050 ± 0.022), but the differences are neither large nor statistically significant (t-tests, $P = 0.085$ and 0.105 , respectively). This suggests that these loci supporting non-consensus topologies did not arise by horizontal transmission from deeply divergent *Wolbachia*. Our allele-divergence analyses in the next subsection further suggest relatively few examples of horizontally acquired alleles among our filtered single-copy transcribed loci that comprise bifurcating genomic regions.

From the 7 *Wolbachia* from filarial nematodes, we extracted 143 single-copy genes (133,677 bp). Of these, 83 produced fully resolved single gene trees with posterior probabilities >0.95 at each node. All 83 genes produced the same phylogeny, which recapitulated the host phylogeny: (((*B. malayi*, *B. pahangi*), *W. bancrofti*), *L. sigmodontis*), ((*O. volvulus*, *O. ochengi*), *D. immitis*)). Like

our recombination analyses above, this analysis supported the bifurcating evolution of the filarial nematode *Wolbachia* over tens of millions of years.

Analyzing polymorphism and allelic differences at individual loci. If horizontal acquisition of single-copy genes from distantly related *Wolbachia* occurs frequently (e.g. on a timescale of hundreds or a few thousand years), outlier alleles should be most apparent in sets of closely related genomes. To assess the timing and pervasiveness of horizontal gene transfer among *Wolbachia* genomes, we examined in turn genome sets showing 3 levels of differentiation. First, we considered representatives of the wRi-like *Wolbachia* from 8 *Drosophila* hosts (Turelli et al. 2018). For the wRi-like set, the 525 single-copy genes showed a maximum observed pairwise third-position divergence (estimated number of substitutions per site) of 2.4×10^{-4} (wAna vs wTri) after Jukes-Cantor correction. Over the 438 single-copy genes, the wMel-like variants differed by at most 2.1×10^{-3} (wRec vs wZts). Finally, over 442 single-copy genes, the 7 supergroup A genomes showed a maximum third-position divergence of 3.1×10^{-2} (wBic vs wTri). Hence, our 3 sets span 2 orders of magnitude in their differentiation.

Based on these pairwise differences, we applied a heuristic criterion for identifying plausible examples of horizontally acquired alleles. For individual loci within a specific set of *Wolbachia*, we categorized an allele as a “2-fold outlier” if it differed from all other alleles in the set at twice as many nucleotides as any of the remaining alleles differed from one another. This heuristic criterion is obviously most easily satisfied for the least diverged wRi-like clade. However, across the 525 single-copy gene set, only one relatively long single-copy gene (1,915 bp) showed 4 variable nucleotide sites; the remainder showed 3 or fewer. Thus, there were no 2-fold outlier loci. As noted above, recombination between wRi-like variants would have been essentially impossible to detect, but even if it occurred, it would not have significantly distorted the bifurcating divergence of genomic regions composed of single-copy loci on which we based our chronograms (see the following Results sections). However, our criteria for selecting loci to include in these analyses, namely that they be single-copy and equal length for at least $n-1$ of the n alleles compared, may have screened out some loci with evidence of horizontally acquired alleles.

To identify outlier alleles in the wMel-like and 7 supergroup A gene sets, we ranked loci in each set by the fraction of variable nucleotides, under the assumption that the most variable loci are the most promising candidates for horizontal transmission. Among the 20 wMel-like loci with the largest fraction of variable nucleotides, only 4 showed a 2-fold outlier allele (Supplementary Table 6), and only the second most-variable gene initially appeared as a plausible candidate for horizontal acquisition of distantly related alleles (Supplementary Information). This gene is homologous to RefSeq_WP_010962738.1, peptidase M48, with 43 variable sites across the 1,282 bp (3.35%). The wBor allele differs from the other 19 alleles at 34 to 36 sites (2.7% to 2.8%), while the other alleles differ from one another at fewer than 5 sites ($\leq 0.39\%$). The wBor allele appears intact and ends with an in-frame stop codon. A 55 bp segment at the 3' end of the wBor allele matches a wMel ISWpi1 insertion sequence element (EU714520.1), with 100% identity on the reverse strand. There are many ISWpi1 elements in the wBor genome, and one outside the peptidase M48 (on a different contig) is an exact match of the one within peptidase M48. All other 19 alleles are identical across this 55 bp segment, each differs from the wBor allele at 34 sites. On the 5' end of this gene in wBor, the next single-copy gene is homologous to C4-dihydroxy-2-butanone-4-phosphate synthase, which shows less than 1% divergence between wBor and wMel. On the 3' end,

there are 2 transposases followed by a membrane protein in wBor, while the transposases are absent and the membrane protein is present in wMel.

BLAST found that the wBor allele shares 99.9% similarity with an allele in wInn from *D. innubila*. Both 5' and 3' flanking regions of the allele in wInn match those in wBor, including the 2 transposases at the 3' end. The transposases closer to peptidase M48 in wBor and wInn are identical. Taken together, these data are consistent with IS activity within the genome of the wInn-wBor MRCA generating the 55 bp fragment that distinguishes them from the 19 other alleles in our analysis. While transposon insertions from the same genome affect terminal branch lengths in a phylogram, they do not affect topology in the case of our 20 alleles at this gene, given that the 19 other alleles are essentially identical.

In contrast to the wMel-like gene set, we did find limited evidence for horizontal transfer from a distantly related *Wolbachia* in the 7 supergroup A set, with the most variable locus meeting our 2-fold outlier criteria: the gene homologous to RefSeq_WP_010962975.1, a membrane protein. This locus showed about 19% polymorphic sites, with 95 variable sites out of 498 bp. The wHa allele shows 92 to 94 bp differences from the homologous alleles in the other 6 *Wolbachia* (18.5% to 18.9%), whereas the other 6 alleles have a maximum of 3 pairwise differences ($\leq 0.6\%$). BLAST revealed that the wHa copy is identical to copies in *Wolbachia* from diverse host species, including sawflies, hoverflies, moths, carpet beetles, and Tachinidae flies (e.g. Accessions: OZ034713.1, OX366332.1, AP028948.1, OX366359.1, and OZ034764.1). The genes on either side did not seem exceptional. tRNA-Leu is on the 5' side and identical in all 7 genomes. On the 3' side is uroporphyrinogen III methyltransferase (708 bp), where the wHa allele has up to 27 pairwise differences over 708 bp from the other 6, but the other alleles differed from one another at up to 18 bp (less than 2-fold). No other locus met our 2-fold outlier criteria. As noted previously, the 38 loci that supported non-consensus phylogenies did not show significantly more pairwise differences among the alleles than did the loci that supported the consensus phylogeny.

Our gene sets contained most of the single-copy genes in these genomes. In the 1.45 Mb wRi reference genome, Prokka found 1,397 genes, of which 726 had a single copy. Removing what are clearly phage genes and transposases, the number of single-copy genes dropped to 694. Hence, ignoring our alignment and gap criteria, 694 genes (706,146 bp) set the theoretical maximum. The 525 genes we analyzed for “outliers” represent 75% of this theoretical maximum and roughly one-third of the entire wRi genome. In the 1.27 Mb wMel reference genome, Prokka found 1,271 genes, of which 710 were single-copy genes and not phage genes or transposases. The wMel-like and 7 supergroup A gene sets we analyzed included more than half of these and represented about one-third of the entire wMel genome.

It appears that a majority of single-copy loci in our analyses did not acquire exogenous alleles over the timescales of differentiation for our wRi-like and wMel-like variants. Hence, these genomes contain substantial genomic regions composed of single-copy loci that undergo bifurcating evolution from which we can estimate chronograms, once we establish calibrations of *Wolbachia* evolution. In our Supplementary Information, we provide additional details and support for bifurcation.

Wolbachia substitution-rate calibrations

While multiple approaches for calibrating *Wolbachia* divergence times have been developed, they may be unreliable for different reasons. For example, the earliest calibration probably underestimated divergence times because of relatively slow *ftsZ* evolution

Table 3. Absolute and relative divergence of *Wolbachia* vs host nuclear and mitochondrial (mtDNA) genomes for cases of putative cladogenic *Wolbachia* transmission.

Host Species	Wolbachia	Nuclear	mtDNA	Wolbachia/nuclear	mtDNA/Wolbachia	mtDNA/nuclear
<i>Nasonia longicornis</i> (wNlonB1)/ <i>N. giraulti</i> (wNgirB) ¹	0.2, 0.3 (4,486 bp)	NA, 1.12 (4,135 bp)	NA, 45.24 (2,241 bp)	NA, 0.27	NA, 150.8	NA, 40.4
<i>Onchocerca ochengi</i> / <i>O. volvulus</i>	0.55, 0.84 (379,047 bp) ^{2,3}	0.53, 0.77 (33,099 bp) ⁴	7.60, 11.23 (10,373 bp) ⁴	1.04, 1.09	13.8, 13.4	14.3, 14.6
<i>Dirofilaria immitis</i> ⁵ / <i>O. ochengi</i>	23.53, 35.27 (379,047 bp) ³	21.93, 33.42 (33,099 bp) ^{4,5}		1.07, 1.06		
<i>D. immitis</i> / <i>O. volvulus</i>	23.56, 35.33 (379,047 bp) ⁵	22.02, 33.59 (33,099 bp) ^{4,5}		1.07, 1.05		
<i>Brugia malayi</i> ² / <i>B. pahangi</i> ³	0.88, 1.24 (499,230 bp) ⁸	1.94, 2.87 (33,099 bp) ^{6,7}	28.2, 43.7 (10,361 bp) ^{6,7}	0.45, 0.43	32.0, 35.2	14.5, 15.2
<i>Wuchereria bancrofti</i> ⁹ / <i>B. malayi</i>	4.30, 6.10 (499,230 bp) ⁸	7.94, 11.66 (33,099 bp)		0.54, 0.54		
<i>W. bancrofti</i> / <i>B. pahangi</i>	4.30, 6.10 (499,230 bp) ⁸	7.93, 11.48 (33,099 bp)		0.54, 0.53		
Nomada clade: (<i>N. ferruginata</i>, (<i>N. panzeri</i>, (<i>N. flava</i>, <i>N. leucophthalma</i>)))¹⁰						
<i>N. ferruginata</i> / <i>N. panzeri</i>	0.29, 0.46 (613,605 bp)	1.73, 2.39 (36,402 bp)	2.53, 4.95 (9,249 bp)	0.17, 0.19	8.72, 10.76	1.46, 2.07
<i>N. ferruginata</i> / <i>N. flava</i>	0.27, 0.43 (613,605 bp)	1.43, 2.15 (36,402 bp)	2.61, 5.27 (9,249 bp)	0.19, 0.20	9.67, 12.26	1.83, 2.45
<i>N. ferruginata</i> / <i>N. leucophthalma</i>	0.27, 0.43 (613,605 bp)	1.47, 2.13 (36,402 bp)	2.22, 4.45 (9,249 bp)	0.18, 0.20	8.22, 10.35	1.51, 2.09
<i>N. panzeri</i> / <i>N. flava</i>	0.032, 0.051 (613,605 bp)	0.99, 1.29 (36,402 bp)	2.42, 4.94 (9,249 bp)	0.032, 0.040	75.6, 96.9	2.44, 3.83
<i>N. panzeri</i> / <i>N. leucophthalma</i>	0.033, 0.051 (613,605 bp)	1.00, 1.27 (36,402 bp)	2.25, 4.52 (9,249 bp)	0.033, 0.040	68.2, 88.6	2.25, 3.56
<i>N. flava</i> / <i>N. leucophthalma</i>	0.0088, 0.011 (613,605 bp)	0.65, 1.06 (36,402 bp)	1.34, 2.62 (9,249 bp)	0.014, 0.010	152.3, 238.2	2.06, 2.47

For each entry in the body of the table, the first value is the estimated percent substitutions (after Jukes-Cantor correction) per third-position codon site per year, the second value is the estimated percentage of synonymous substitutions across the first and third positions. The values in parentheses are total number of nucleotides in the exons used in the divergence estimates.

Data sources and notes: 1. Raychoudhury et al. (2009). 2. Darby et al. (2012). 3. Data from 387 protein-coding loci. For the *Onchocerca* pair, 572,142 bp of *Wolbachia* data are available. For consistency, the data were trimmed to the loci for which *Wolbachia* sequences from *Dirofilaria immitis* sequences were also available. 4. Nuclear and mtDNA genome data for the *Onchocerca* species and *Wolbachia* data for *Dirofilaria immitis* from the 50 Helminth Genomes Project: <https://www.sanger.ac.uk/collaboration/50hgp/>. 5. Gomes-de-Sa et al. (2022). 6. Ghedin et al. (2007). 7. Lau et al. (2015). 8. Data from 511 transcribed loci. For the *Brugia* pair, 598,257 bp of *Wolbachia* data are available. For consistency, the data were trimmed to the loci for which *Wolbachia* sequences from *Wuchereria bancrofti* were also available. 9. Small et al. (2019). 10. Gerth and Bleidorn (2016).

(Werren et al. 1995; Supplementary Table 7; Supplementary Information), calibrations based on facultative associations may be unreliable if they do not actually involve codivergence (Raychoudhury et al. 2009), and mutation-based calibrations may be biased if mutation rates do not accurately reflect substitution rates (Richardson et al. 2012). We calibrated rates of *Wolbachia* molecular evolution using new data from trios of filarial nematodes (Qing et al. 2025) that are known to codiverge with their obligate *Wolbachia* that they require for survival (Comandatore et al. 2013; Manoj et al. 2021). Table 3 provides estimates for the relative rates of divergence for the *Wolbachia* vs the host nuclear and mitochondrial genomes, derived from 2 trios: ((*O. volvulus*, *O. ochengi*), *D. immitis*) and ((*B. malayi*, *B. pahangi*), *W. bancrofti*). Notably, the relative rates of *Wolbachia* and host nuclear divergence are quite consistent within each trio. However, the ratios for *Wolbachia*/nuclear divergence in the trio ((*B. malayi*, *B. pahangi*), *W. bancrofti*) are only about half that observed in ((*O. volvulus*, *O. ochengi*), *D. immitis*). This underlies the differences in estimated *Wolbachia* substitution rates derived from these alternative trios (Table 4).

We used the support intervals for the *Dirofilaria* and *Onchocerca* divergence time (95% HPD CI of 37.56 to 91.96 MYA) and for the *Brugia* and *Wuchereria* (95% HPD CI of 12.8 to 48.59 MYA) MRCA's estimated by Qing et al. (2025) to approximate support intervals for the substitution rates of the codiverging *Wolbachia*, averaging the differences given in Table 3 between each congener and its respective outgroup. Substitution rates were estimated by dividing the estimated number of substitutions between lineages (after Jukes-Cantor correction) by these divergence time estimates. The *Dirofilaria-Onchocerca* point estimate for the substitution rate per third site per year, 1.82×10^{-9} [with

support interval $(1.28 \text{ to } 3.13) \times 10^{-9}$], is about 2.5 times the *Brugia-Wuchereria* point estimate, 7.42×10^{-10} [with support interval $(4.42 \text{ to } 16.8) \times 10^{-10}$]. However, the support intervals overlap, and the *Dirofilaria-Onchocerca* estimate is consistent with our revised estimate from the Raychoudhury et al. (2009) *Nasonia* example (see below).

Two additional rate estimates were available from the pairs of congeners. As noted above, we used relative chronograms for each host trio to estimate the MRCA ages for each pair. We used the point estimates for the MRCA ages of each trio, together with the support intervals for the relative ages of the congeners, to approximate support intervals for the MRCA ages of the congeners. The resulting point estimates and support intervals for the *Wolbachia* substitution rates appear in Table 4. In both cases, the congener-based rate estimates are consistent with the respective estimates based on comparisons to the outgroup. The range of the 4-point estimates is 6.47×10^{-10} to 1.82×10^{-9} . It is notable that according to Qing et al. (2025), the age of the MRCA of these nematode trios is over 40 MYA, yet the *Wolbachia* substitution-rate estimates vary by only about a factor of 3.

For each of our rate estimates, the support range extends from roughly half to roughly double the point estimate. When estimating chronograms, we approximated this uncertainty by multiplying the point estimate by a lognormal random variable, X , defined by

$$X = \text{Exp}[\text{Normal}(\mu = 0, \sigma = 0.353654)].$$

Setting the mean of the underlying normal to 0 produces a median for X of 1. The value for σ is chosen so that $P[X < 0.5] = P[X > 2] = 0.025$, approximating our empirical support intervals for the rates. The value

Table 4. Substitution rate estimates used to calibrate *Wolbachia* chronograms.

Method	Substitutions/site/year			Data	Time calibration/validation
	Third sites	Synonymous sites			
"Universal" estimate of the average substitution rate for coding DNA in bacteria applied to <i>Wolbachia</i> <i>ftsZ</i> ¹	NA	7 to 8×10^{-9}		<i>ftsZ</i> differences among <i>Wolbachia</i> ¹	Divergence-time estimates based on 21 kb of coding DNA from bacteria <i>Salmonella typhimurium</i> and <i>Escherichia coli</i> ² supplemented with data indicating similar synonymous-site substitution rates for <i>ftsZ</i> in <i>Wolbachia</i> ¹ .
Divergence of mitochondrial vs <i>Wolbachia</i> genomes among <i>Drosophila melanogaster</i> isofemale lines ³	6.87×10^{-9}	NA		mtDNA and <i>Wolbachia</i> genomes	Compare sequence differences among isofemale lines for mtDNA and <i>Wolbachia</i> . Calibrate <i>Wolbachia</i> substitution rates using the per-generation mtDNA mutation rate as a prior for "short term" mtDNA (and <i>Wolbachia</i>) evolution ³ .
Plausible example of codivergence of specific <i>Wolbachia</i> variants in 2 <i>Nasonia</i> species Updated universal synonymous-site substitution rate ⁴ for bacteria cross-validated with a host nuclear clock ⁵ for synonymous sites in <i>Wolbachia</i> (<i>wNlonB1</i> vs <i>wNgrB</i>) putatively cladogenically transmitted between <i>Nasonia longicornis</i> and <i>N. giraulti</i>	2.2×10^{-9}	3.3×10^{-9}		Fragments of <i>Wolbachia</i> and host nuclear loci	Concordance of divergence times estimated from molecular clocks applied to <i>Wolbachia</i> and host nuclear DNA ⁵ . Our substitution-rate estimates assume MRCA for host and <i>Wolbachia</i> 0.46 MYA, averaging the bacterial and eukaryotic clock-based estimates from Raychoudhury et al. (2009).
Codivergence of obligate symbiotic <i>Wolbachia</i> with filarial nematodes <i>Onchocerca volvulus</i> / <i>O. ochengi</i>	1.38×10^{-9} (0.73 to 3.44×10^{-9})	2.11×10^{-9} (1.11 to 5.25×10^{-9})	$\times$	Draft <i>Wolbachia</i> and host nuclear genomes	Time back to MRCA of hosts (and 95% credible interval), 2.01 (0.80 to 3.79) MY, estimated from our relaxed-clock relative chronogram (Fig. 2b) for <i>Onchocerca</i> and <i>Dirofilaria immitis</i> , assuming 64.73 MYA as age of the MRCA of <i>Onchocerca volvulus</i> and <i>Dirofilaria immitis</i> . The fossil-calibrated estimate (and support interval) for the age of the MRCA for <i>Onchocerca volvulus</i> and <i>Dirofilaria immitis</i> is 64.73 (37.56 to 91.96) MYA ⁶ . Rates estimated by averaging the <i>Wolbachia</i> divergences from Table 3 between <i>Dirofilaria immitis</i> and the two <i>Onchocerca</i> species.
<i>Dirofilaria immitis</i> / <i>Onchocerca</i>	1.82×10^{-9} (1.28 to 3.13×10^{-9})	2.73×10^{-9} (1.92 to 4.70×10^{-9})	$\times$	Draft <i>Wolbachia</i> and host nuclear genomes	
<i>Brugia malayi</i> / <i>B. pahangi</i>	6.47×10^{-10} (3.71 to 16.4×10^{-10})	9.12×10^{-10} (5.23 to 23.1×10^{-10})		Draft <i>Wolbachia</i> and host nuclear genomes	Time back to MRCA of <i>Brugia</i> hosts (and 95% credible interval), 6.66 (2.68 to 11.85) MY, from our relaxed-clock relative chronogram (Fig. 2c) for <i>Brugia</i> and <i>Wuchereria bancrofti</i> , assuming 28.97 MYA as age of the MRCA of <i>Brugia malayi</i> and <i>Wuchereria bancrofti</i> ⁵ . Support intervals for rates from support intervals for MRCA age.
<i>Wuchereria bancrofti</i> / <i>Brugia</i>	7.42×10^{-10} (4.42 to 16.8×10^{-10})	1.05×10^{-9} (6.28 to 23.8×10^{-10})		Draft <i>Wolbachia</i> and host nuclear genomes	The fossil-calibrated estimate (and support interval) for the age of the MRCA for <i>Brugia malayi</i> and <i>Wuchereria bancrofti</i> is 28.97 (12.8 to 48.59) MYA ⁶ . Rates estimated by averaging <i>Wolbachia</i> divergences from Table 3 between <i>W. bancrofti</i> and the two <i>Brugia</i> species.
Three alternative <i>Wolbachia</i> calibrations based on the <i>Nomada</i> bee clade: (<i>N. ferruginata</i>, (<i>N. panzeri</i>, (<i>N. flava</i>, <i>N. leucophthalma</i>)) <i>Wolbachia</i> in <i>N. ferruginata</i> vs 3-species ingroup	6.46×10^{-10}	1.03×10^{-9}		Draft <i>Wolbachia</i> and host nuclear genomes ⁷	Estimated time back to MRCA of hosts, 2.14 MY (Gerth and Bleidorn (2016, p. 4). Rates estimated by averaging the 3 sequence divergence estimates in Table 3.
<i>Wolbachia</i> in <i>N. panzeri</i> vs (<i>N. flava</i> , <i>N. leucophthalma</i>) (95% credible intervals based on approximate credible range of divergence times)	1.31×10^{-10} (0.84 to 2.29×10^{-10})	2.06×10^{-10} (1.32 to 3.59×10^{-10})		Draft <i>Wolbachia</i> and host nuclear genomes	Time back to MRCA of hosts and 95% credible interval, 1.24 (0.71 to 1.93) MY, obtained from our relative chronogram (Fig. 2a), assuming a crown age of 2.14 MY for the 4-species clade. The rates were estimated by averaging the two sequence divergence estimates in Table 3.
<i>Wolbachia</i> in <i>N. flava</i> vs <i>N. leucophthalma</i> (95% credible intervals based on approximate credible range of divergence times)	5.71×10^{-11} (3.36 to 12.9×10^{-11})	7.14×10^{-11} (4.20 to 16.2×10^{-11})		Draft <i>Wolbachia</i> and host nuclear genomes	Time back to MRCA of hosts and 95% credible interval, 0.77 (0.34 to 1.31) MY, obtained from our relative chronogram (Fig. 2a), assuming a crown age of 2.14 MY for the 4-species clade.

References. 1. Werren et al. (1995). 2. Ochman and Wilson (1987). 3. Richardson et al. (2012). 4. Ochman et al. (1999). 5. Raychoudhury et al. (2009). 6. Qing et al. (2025). 7. Gerth and Bleidorn (2016).

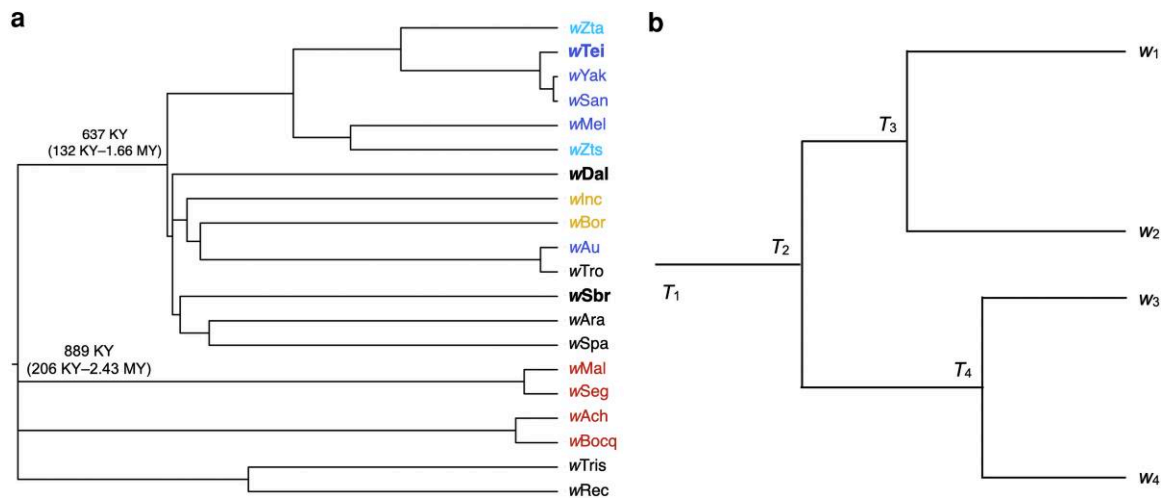


Fig. 4. a) The timescale of *wMel*-like *Wolbachia* movements and limitations of inferences. An absolute chronogram for the *wMel*-like *Wolbachia* in our study. Labels for the *Wolbachia* associated with the most distantly related hosts presented in Fig. 1b are in bold and include: *wDal* (in *D. alloeum*), *wSbr* (in *S. brevicornis*) and *wTei* (in *D. teissieri*) diverged 637 KY, with a CPR of 132 KY to 1.66 MY. The colored *Wolbachia* labels match the host clades presented in Fig. 1b. The crown age is 889 KY, with CPR of 206 KY to 2.43 MY. The several polytomies reflect the rapid periods of diversification noted in Fig. 1a. b) An illustration clarifying the limitations of inferences from a chronogram based on single *Wolbachia* genomes from individual hosts. These limitations do not affect our central inferences concerning the timing of *Wolbachia* movement and molecular evolution.

for σ implies that $E[X] = 1.06$, so that the mean rate is close to the point estimate.

To produce plausible estimates of divergence times and conservative estimates of their biological and statistical uncertainty, we averaged the fastest and slowest of our nematode-based rates and then constructed a “conservative plausible range” (CPR) for each estimate using the statistical lower bound from the fastest rate and the upper bound from the slowest rate. This CPR more realistically captures the biological and statistical uncertainty of our time estimates than would conventional support intervals based on the statistical uncertainty of a single calibration. In Table 4, we present these new rate estimates as well as older estimates based on the mutation-based *Drosophila melanogaster* estimate from Richardson et al. (2012), a *Nasonia* calibration based on plausible codivergence of a specific *Wolbachia* lineage between 2 *Nasonia* species, and 3 alternative rates estimated from *Nomada* data of Gerth and Bleidorn (2016). With firm evidence for codivergence of *Wolbachia* and hosts in the 2 trios of filarial nematodes, and fossil-based estimates of the nematode host divergence times, we believe they are more reliable than previous estimates from *Nasonia* and *Nomada*. Notably, the fastest *Nomada* rate is consistent with the slowest of our nematode-based calibrations, and the *Nasonia* rate is consistent with the fastest of our nematode-based calibrations. However, as shown in Table 4, there are 2 additional *Nomada* rate estimates, based on the ingroup species (*N. panzeri*, *N. flava*, and *N. leucophthalma*). Those estimates are much slower than any of our nematode estimates. These slower rates seem to be artifacts of horizontal *Wolbachia* acquisition in these 3 ingroup *Nomada*. The key observations are that (i) the most closely related *Nomada* hosts yield 10-fold slower *Wolbachia* rates, while (ii) the estimated mtDNA and nuclear rates are relatively constant. Rather than invoking a recent 10-fold slowdown of *Wolbachia* evolution over the last million years, these data seem more plausibly explained by horizontal *Wolbachia* acquisition within the 3-species *Nomada* ingroup.

For our *Wolbachia* chronograms, we focus on age estimates and support intervals derived from the nematodes. The alternative calibrations are considered further in our Discussion and Supplementary Information. Based on our new rate estimates

(Table 4), which were consistent with estimates from *Nasonia* and *Nomada*, the Richardson et al. (2012) rate seems too fast by about a factor of 7. Turelli et al. (2018) used the Richardson et al. (2012) rate. Using our nematode-based calibrations, we now estimate *wRi*-*wSuz* divergence as 76 KYA (CPR: 14 to 218 KYA). The implications of this revision are considered in our Discussion.

Applying nematode-based calibrations to *wMel*-like *Wolbachia*

Estimating *wMel*-like *Wolbachia* divergence times. Using our new nematode-based calibrations, we estimated the timescale of *wMel*-like *Wolbachia* host switching. Figure 4a presents an estimated *Wolbachia* chronogram for the variants presented in Fig. 1a. Using bifurcating genomic regions with little evidence of horizontal allele acquisition, these *Wolbachia* genomes diverged only about 206 KY to 2.43 MYA (Fig. 4a, Table 5). The *wMel*-like *Wolbachia* (*wDal*, (*wSbr*, *wTei*)) in the most-diverged hosts (*Diachasma alloeum*, (*S. brevicornis*, *D. teissieri*)) diverged about 637 KYA, with a CPR of 132 KYA to 1.7 MYA, that is, over a time-scale comparable to a single host speciation event (Coyne and Orr 2004). The estimate for *wMel*-like divergence is based on the *wDal*-*wSYT* (i.e. the clade including *wSan*, *wYak*, and *wTei*) node in Fig. 4a, which is the smallest clade that includes *wDal*, *wSbr*, and *wTei*. Among *wMel*-like *Wolbachia* observed in 18 drosophilids in our study (Fig. 1), *wZts* in tropical *Zaprionus tsacasi* is the most closely related to *wMel* in *D. melanogaster*, diverging about 336 KYA (CPR: 67 to 924 KYA). We are unaware of any other *Wolbachia* more closely related to *wMel* (Supplementary Information, Supplementary Fig. 1), although we expect others will be discovered as additional *Wolbachia* are sequenced. *Zaprionus* flies are members of the (paraphyletic) *Drosophila* subgenus that diverged from the *Sophophora* subgenus about 47 MYA (95% CI: 43.9 to 49.9 MYA), highlighting proliferation of *wMel*-like *Wolbachia* across species that span the entire paraphyletic genus also named *Drosophila* (Suvorov et al. 2022).

Determining *wMel*-like *Wolbachia* acquisition modes. Many obligate mutualistic endosymbionts like the *Wolbachia* in filarial nematodes (Comandatore et al. 2013) and *Buchnera* in aphids (Baumann et al. 1995) are acquired cladogenically. In contrast,

Table 5. wMel-like Wolbachia node age point estimates and 95% CIs in years for all nodes presented in Fig. 4a.

Node	Drosophila mtDNA			Dirofilaria (fixed)			Dirofilaria (lognormal)			Brugia (fixed)			Brugia (lognormal)			Plausible Time Estimates		
	Point	95% CI	Point	95% CI	Point	95% CI	Point	95% CI	Point	95% CI	Point	95% CI	Point	95% CI	Point	Point	CPR	CPR
Root	122,104	(98,666, 152,513)	460,910	(372,437, 575,693)	466,869	(206,051, 900,601)	1,296,533	(1,047,660, 1,619,410)	1,311,857	(618,461, 2,433,090)	889,363	(206,051, 2,433,090)	889,363	(618,461, 2,433,090)	889,363	(206,051, 2,433,090)		
(wSan,wYak)	877	(157, 2,248)	3,311	(593, 8,488)	3,330	(496, 10,133)	9,314	(1,670, 23,878)	9,498	(971, 28,857)	6,414	(496, 28,857)	6,414	(971, 28,857)	6,414	(496, 28,857)		
(wTei,(wSan,wYak))	3,904	(1,760, 7,043)	14,739	(6,645, 26,588)	14,777	(4,948, 34,677)	41,461	(18,692, 74,791)	42,241	(12,162, 93,988)	28,509	(4,948, 93,988)	28,509	(12,162, 93,988)	28,509	(4,948, 93,988)		
(wZta,(wTei,(wSan,wYak)))	35,452	(22,890, 48,999)	133,824	(86,406, 184,959)	134,736	(46,652, 266,041)	376,446	(243,059, 520,287)	376,929	(130,538, 734,069)	255,832.5	(46,652, 734,069)	255,832.5	(130,538, 734,069)	255,832.5	(46,652, 734,069)		
(wMel,wZts)	46,829	(35,337, 61,459)	176,766	(133,390, 231,991)	177,819	(66,835, 333,151)	497,242	(375,224, 652,588)	494,263	(195,489, 924,345)	336,041	(66,835, 924,345)	336,041	(195,489, 924,345)	336,041	(66,835, 924,345)		
((wMel,wZts),(wZta,wSYT))	59,793	(48,403, 73,030)	225,702	(182,710, 275,670)	227,964	(96,325, 434,420)	634,896	(513,960, 775,454)	640,685	(268,781, 1,161,800)	434,324.5	(96,325, 1,161,800)	434,324.5	(268,781, 1,161,800)	434,324.5	(96,325, 1,161,800)		
(wTro,wAu)	3,796	(1,395, 7,104)	14,332	(5,267, 26,818)	14,860	(4,106, 35,845)	40,315	(14,818, 75,439)	41,231	(10,924, 93,382)	28,045.5	(4,106, 93,382)	28,045.5	(10,924, 93,382)	28,045.5	(4,106, 93,382)		
(wBor,(wTro,wAu))	80,825	(68,085, 94,318)	305,093	(257,003, 356,026)	309,441	(134,625, 595,353)	858,222	(722,944, 1,001,500)	866,659	(370,028, 1,562,040)	588,050	(134,625, 1,562,040)	588,050	(370,028, 1,562,040)	588,050	(134,625, 1,562,040)		
(wInc,(wBor,(wTro,wAu)))	83,855	(71,365, 96,353)	316,530	(269,383, 363,707)	320,611	(125,589, 601,643)	890,395	(757,771, 1,023,100)	900,733	(387,630, 1,600,610)	610,672	(125,589, 1,600,610)	610,672	(387,630, 1,600,610)	610,672	(125,589, 1,600,610)		
(wAra,wSpa)	78,828	(60,822, 95,936)	297,556	(229,588, 362,133)	297,397	(126,652, 584,033)	837,021	(645,826, 1,018,670)	842,347	(312,464, 1,521,920)	569,872	(126,652, 1,521,920)	569,872	(312,464, 1,521,920)	569,872	(126,652, 1,521,920)		
(wSbr,(wAra,wSpa))	85,407	(72,389, 94,497)	322,389	(273,248, 371,800)	324,318	(134,375, 617,210)	906,875	(768,643, 1,045,870)	919,205	(371,034, 1,620,360)	621,761.5	(134,375, 1,620,360)	621,761.5	(371,034, 1,620,360)	621,761.5	(134,375, 1,620,360)		
(wDal,(wAra-wSbr,wTro-wInc))	87,120	(75,334, 100,665)	328,856	(284,368, 379,984)	332,375	(132,290, 622,578)	925,068	(799,922, 1,068,890)	940,738	(394,476, 1,658,090)	636,556.5	(132,290, 1,658,090)	636,556.5	(394,476, 1,658,090)	636,556.5	(132,290, 1,658,090)		
wSYT-wDal	88,309	(77,169, 101,950)	333,344	(291,294, 384,833)	337,106	(134,946, 627,989)	937,691	(819,406, 1,082,530)	949,467	(401,110, 1,684,930)	643,286.5	(134,946, 1,684,930)	643,286.5	(401,110, 1,684,930)	643,286.5	(134,946, 1,684,930)		
(wMal,wSeg)	7,496	(3,328, 12,777)	28,297	(12,564, 48,230)	29,227	(7,380, 62,302)	79,600	(35,343, 135,672)	81,049	(26,366, 174,569)	55,138	(7,380, 174,569)	55,138	(26,366, 174,569)	55,138	(7,380, 174,569)		
(wBocq,wAch)	9,404	(4,754, 15,762)	35,500	(17,948, 59,497)	36,314	(11,089, 76,932)	99,861	(50,488, 167,354)	100,637	(29,261, 221,975)	68,475.5	(11,089, 221,975)	68,475.5	(29,261, 221,975)	68,475.5	(11,089, 221,975)		
(wTris,wRec)	69,956	(44,884, 113,664)	264,067	(169,425, 429,051)	272,767	(86,711, 565,736)	742,817	(476,590, 1,206,910)	764,425	(246,849, 1,566,220)	518,596	(86,711, 1,566,220)	518,596	(246,849, 1,566,220)	518,596	(86,711, 1,566,220)		

We present estimates produced using 5 different calibrations described in the Materials and Methods and Results sections. The final columns provide our plausible point estimates calculated as the average of the point estimates of the fastest (Dirofilaria, lognormal) and slowest (Brugia, lognormal) nematode-based calibrations. We report the conservative plausible range (CPR), which ranges from the lower-bound of the fastest and the upper-bound of the slowest nematode-based calibrations. As described in the text, the range of our new nematode-based calibrations are consistent with estimates from *Nasonia* and *Nomada*. In contrast, the mutation-based rate estimate of Richardson et al. (2012) is significantly faster (Table 4).

because the divergence times for our *wMel*-like *Wolbachia* are so much shorter than for their hosts, all but one of these variants must have been acquired through introgression or non-sexual horizontal transfer (although as explained in the next subsection, we cannot exclude that one host cladogenically acquired its *Wolbachia* from a closely related host we did not sample). Introgression is plausible only between the most closely related drosophilid species in our study (i.e. within the colored triangles in Fig. 1b), some of which may hybridize (Matute and Cooper 2021). Confirming introgressive acquisition requires demonstrating concordant divergence times for mtDNA and the associated *Wolbachia*, and these must be more recent than host divergence.

Joint analysis of mtDNA and *Wolbachia* sequence divergence in Cooper et al. (2019) implied that the 3-species *D. yakuba* clade (*D. teissieri*, (*D. yakuba*, *D. santomea*)) first acquired *Wolbachia* by horizontal transfer from an unknown host. The *Wolbachia* were then transferred within the clade through hybridization and introgression. *D. yakuba* hybridizes with its endemic sister species *D. santomea* on the island of São Tomé (Lachaise et al. 2000; Comeault et al. 2016; Cooper et al. 2017), and with *D. teissieri* on the edges of forests on the nearby island of Bioko (Cooper et al. 2018). *Wolbachia* and mtDNA chronograms were generally concordant for these 3 hosts and indicated more recent common ancestry for these maternally inherited cytoplasmic factors than for the bulk of their hosts' nuclear genomes (Cooper et al. 2019).

wZta from *Z. taronus* is the most closely related *Wolbachia* to *wSYT*, diverging about 256 KYA (CPR: 47 to 734 KYA). *Z. taronus* also occurs on São Tomé, particularly co-occurring with *D. santomea* at high altitudes on Pico de São Tomé; but it diverged from the *D. yakuba* triad about 47 MYA (Fig. 1b), making introgression between *Z. taronus* and the *D. yakuba*-clade species impossible (cf. Coyne and Orr 2004). Yet, its *Wolbachia* (*wZta*) diverged from the *wYak*-clade *Wolbachia* only about 47 to 734 KYA (Fig. 4a). These data illustrate horizontal *Wolbachia* transfer between distantly related species with overlapping ranges and habitats. The mechanisms of horizontal transfer are largely unresolved, but plausible alternatives are considered in our Discussion.

To identify possible cases of introgressive *Wolbachia* transfer, we compared estimated divergence times for mtDNA and *Wolbachia*. Table 6 presents the estimated mtDNA substitution rates per third-position site per year for the 5 *Drosophila* pairs described in our Materials and Methods, comparing rates derived from recently diverged species, which may plausibly hybridize, with the slower rates estimated from more distantly related species unlikely to produce fertile hybrids. To calibrate mtDNA divergence, we averaged the point estimates for mtDNA substitution rates per third-position site per year for *D. melanogaster*-*D. simulans* [1.56×10^{-8} with support interval $(1.28 \text{ to } 1.93) \times 10^{-8}$] and *D. erecta*-*D. yakuba* [1.76×10^{-8} with support interval $(1.50 \text{ to } 2.09) \times 10^{-8}$]. We used the average rate, 1.66×10^{-8} , to estimate mtDNA divergence times for relatively recently diverged *Drosophila*. We considered the fastest and slowest mtDNA substitution rates across both credible intervals as a conservative substitution-rate range (1.28×10^{-8} , 2.09×10^{-8}).

We tested for introgressive *wMel*-like *Wolbachia* transfer between 4 species pairs for which hybridization and introgression may be possible by comparing *Wolbachia* and mtDNA divergence-time estimates with estimates of the hosts' divergence time. We considered 2 relatively closely related species pairs in the *D. montium* subgroup, (*D. seguyi*, *D. malagassya*) and (*D. bocqueti*, *D. sp. aff. chauvacae*) that diverged approximately 2.3 MYA (95% CI: 1.1 to 4.1 MYA) and 2.5 MYA (95% CI: 2.0 to 3.1 MYA), respectively

Table 6. Rates used to estimate mtDNA divergence times.

Species pair	MRCA (Point, MY) ^a	MRCA (95% CI, MY) ^a	mtDNA Length Compared (bp)	mtDNA 3rd-position differences (tip-to-tip) ^b	mtDNA substitutions per 3rd-position site per year (Point)	mtDNA substitutions per 3rd-position site per year (CI)
<i>D. melanogaster</i> - <i>D. simulans</i>	3.62	(2.92 to 4.4)	11,184	11.3%	1.56×10^{-8}	$(1.28 \times 10^{-8}, 1.93 \times 10^{-8})$
<i>D. erecta</i> - <i>D. yakuba</i>	5.47	(4.59 to 6.38)	11,172	19.2%	1.76×10^{-8}	$(1.50 \times 10^{-8}, 2.09 \times 10^{-8})$
<i>D. mojavensis</i> - <i>D. virilis</i>	24.34	(21.94 to 26.88)	11,164	33%	6.78×10^{-9}	$(6.14 \times 10^{-9}, 7.52 \times 10^{-9})$
<i>D. pseudoobscura</i> - <i>D. suzukii</i>	32.78	(29.77 to 35.76)	11,182	29.4%	4.48×10^{-9}	$(4.11 \times 10^{-9}, 4.94 \times 10^{-9})$
<i>D. grimshawi</i> - <i>D. willistoni</i>	46.84	(43.85 to 49.85)	11,184	32.6%	3.48×10^{-9}	$(3.27 \times 10^{-9}, 3.72 \times 10^{-9})$

Five *Drosophila* species with increasing divergence times were used to compute mtDNA 3rd-position differences and estimate mtDNA substitution rates. As described in the main text, we consider the average of the rates derived from the 2 most recently diverged species pairs (i.e. 1.66×10^{-8}) as a plausible point estimate for mtDNA substitutions per third-position site per year between pairs of species that may hybridize. We consider the fastest and slowest substitution rates across the credible intervals for these 2 pairs as a conservative range of substitution rates (1.28×10^{-8} , 2.09×10^{-8}) for approximating mtDNA divergence times for pairs of hosts with *wMel*-like *Wolbachia*. MRCA time estimates and confidence intervals for the hosts are taken from Suvorov et al. (2022). Point estimates and confidence intervals for mtDNA rates are calculated from the observed mtDNA divergences and MRCA time estimates.

^aNuclear MRCA based on the fossil-calibrated chronogram of Suvorov et al. (2022).

^bNucleotide differences were counted and the Jukes-Cantor correction for multiple substitutions was applied.

(see [Materials and Methods](#)). Their mtDNA third-position coding sites differed by 0.53 and 1.15%, respectively. Applying the Jukes-Cantor formula to estimate the number of substitutions and our mtDNA calibration, we estimated that the mtDNA of *D. seguyi* and *D. malagasya* diverged approximately 160 KYA (with a plausible interval of 126 to 207 KYA), suggesting mtDNA introgression. Given our CPR of *wSeg-wMal* *Wolbachia* divergence (7 to 175 KY), we determined that introgressive *Wolbachia* acquisition is possible, but we could not rule out horizontal acquisition after mtDNA introgression. In contrast, the mtDNA of *D. bocqueti* and *D. sp. aff. chaudiacae* diverged approximately 346 KYA (with plausible interval 274 to 449 KYA), while their *wBocq-wAch* *Wolbachia* diverged more recently (CPR: 11 to 222 KY) ([Fig. 4a](#)). Thus, we inferred horizontal *Wolbachia* acquisition.

The more distantly related pairs (*Z. taronus*, *Z. tsacasi*) and (*D. borealis*, *D. incompta*) diverged approximately 14.2 MYA (95% support interval: 12.3 to 16.7 MYA) and 8.2 MYA (95% support interval: 3.7 to 14.5 MYA), respectively. For these pairs, mtDNA third-position coding sites differed by 19.3% and 30%, respectively ([Suvorov et al. 2022](#)), making mtDNA introgression plausible for the *Zaprionus* pair. Applying the mtDNA rate 1.66×10^{-8} after Jukes-Cantor correction, we estimated mtDNA divergence times on the order of 6.7 MYA (with plausible interval 4.6 to 7.5 MYA) for *Z. taronus* and *Z. tsacasi* and 11.5 MYA (with plausible interval 7.1 to 11.7 MYA) for *D. borealis* and *D. incompta*. The latter argued against mtDNA introgression between *D. borealis* and *D. incompta*. In comparison, our plausible divergence interval for their *Wolbachia*, *wBor* and *wInc*, was (126 KY to 1.6 MY), which strongly supported horizontal *Wolbachia* acquisition. The shorter estimated mtDNA divergence time for the *Zaprionus* pair may have reflected a slowdown in mtDNA divergence associated with saturation ([Ho et al. 2005](#)), as illustrated in [Table 6](#) for the 3 more diverged *Drosophila* species pairs. However, irrespective of whether mtDNA introgression occurred in the *Zaprionus* pair, the CPR for *wZta-wZts* divergence (96 KY to 1.2 MY) was much shorter than plausible mtDNA divergence-time estimates, again supporting horizontal *Wolbachia* acquisition.

In summary, we ruled out introgressive *Wolbachia* acquisition for 3 of the 4 possible cases we examined. Introgression was consistent with our analyses only for *wSeg* and *wMal*, but horizontal acquisition could not be confidently rejected.

Identifying limitations of inferences from *Wolbachia* phylogenies and chronograms

Inferences from our chronograms of *wMel*-like variants are limited because we considered only one *Wolbachia* sequence from each host. Specifically, we cannot determine which *Wolbachia* infection was acquired first, whether 2 sister tips on our *Wolbachia* chronograms correspond to one of those hosts acquiring its *Wolbachia* (by either introgression or non-sexual horizontal transmission) from the host of the sister *Wolbachia* variant—or whether one or both *Wolbachia* were recently acquired from an unsampled intermediate host. Nevertheless, our central inferences concerning the timing of *Wolbachia* movement and molecular evolution should be robust. [Turelli et al. \(2018\)](#) previously proposed that *wRi* in *D. simulans* was acquired horizontally from *D. ananassae* because multiple *Wolbachia* sequences were available from each species, and the *wRi* sequences formed a clade nested within a paraphyletic cluster containing the *Wolbachia* from *D. ananassae*. As noted by [Richardson et al. \(2012\)](#), divergence times for extant *Wolbachia* variants in a host species may be much shorter than the duration of its current *Wolbachia* variant because selective

sweeps and genetic drift will eliminate sequence variation that might provide information about the duration of associations.

[Figure 4b](#) illustrates the inferences about *Wolbachia* divergence and acquisition possible from individual *Wolbachia* sequences from various hosts. We considered 4 hypothetical host species with horizontally acquired *Wolbachia*. Host 1 acquired its *Wolbachia*, denoted w_1 , at time T_1 , and we denote by $T_i(j \gg i)$ as the time at which host i acquired its *Wolbachia*, denoted w_i , from host j . Note that T_1 may be much longer than T_2 , T_3 , and T_4 , corresponding possibly to cladogenic inheritance of w_1 . Our chronogram of *wMel*-like variants says nothing about the length of this root branch. If we postulate $T_1 > T_2(1 \gg 2) > T_3(1 \gg 3) > T_4(2 \gg 4)$, the resulting chronogram is [Fig. 4b](#). The chronogram provides no indication that variant w_1 was acquired first (or how it was acquired) or that w_2 was acquired from host 1. Thus, we can infer only that all but one of the variants shown in [Fig. 4a](#) were acquired in the last 206 KY to 2.43 MY and that a transfer between a dipteran and a hymenopteran host occurred within the last 132 KY to 1.66 MY, possibly mediated by intermediate hosts. Moreover, the chronogram [Fig. 4a](#) is consistent with many of these hosts, for which we do not have multiple *Wolbachia* sequences, acquiring their current *Wolbachia* very recently, followed by spatial spread in the new host, as observed for several *Wolbachia*–host associations ([Turelli et al. 2018](#)).

Estimating Wovirus and cif turnover in *wMel*-like and *wRi*-like *Wolbachia*

We next applied our nematode-based calibration to comparisons of closely related *wMel*-like and *wRi*-like *Wolbachia* to understand the timescale of Wovirus and cif turnover occurring outside the bifurcating genomic regions. While *Wolbachia* genome quality varied among our samples, the timescale estimates below are established by focal comparisons involving fully circularized genomes and based on divergence times of the bifurcating genomic regions.

Estimating Wovirus turnover. We used serine recombinase (sr) alleles to type Woviruses in *wMel*-like and *wRi*-like genomes (see our [Materials and Methods](#)), focusing on alleles (sr1 to sr3) of 3 Wovirus Types that are known to contain cifs (sr1WO–sr3WO) ([Bordenstein and Bordenstein 2022](#)). [Supplementary Table 9](#) summarizes our results. While we did not observe sr1WO in any *wMel*-like *Wolbachia*, identical sr1 alleles were observed in all *wRi*-like *Wolbachia*, except for *wAur* and *wTri*. Based on the CPR of *wRi*-like variant divergence, this implies loss of sr1WO by *wAur* and *wTri* in the last 14 to 218 KYA. We observed closely related sr2WO in all *wRi*-like *Wolbachia*, with each sr2 allele differing from the copy in *wRi* at ≤ 2 sites. Among *wMel*-like *Wolbachia*, an sr2WO was observed only in *wSYTZ* (*wSYT* plus the *wZta* out-group) and in *wTris* genomes ([Supplementary Fig. 2a](#)). The absence of this sr2WO in the *wMel* and *wZts* genomes implies gain or loss by *wSYTZ* in the last 96 KYA to 1.16 MYA. The *wSYT* and *wMel* genomes are fully circularized. The absence of this sr2WO in the *wRec* genome, implies gain or loss since its divergence from *wTris* 87 KYA to 1.57 MYA. sr3WO Wovirus was more common, occurring in all *wRi*-like and *wMel*-like genomes, and often in multiple copies ([Supplementary Fig. 2b](#), [Supplementary Table 9](#)). Analysis of intralocus recombination using GARD identified 2 breakpoints in sr3 that we visually confirmed, resulting in 3 partitions. Nodes within partitions were often unresolved and relationships between partitions often varied ([Supplementary Fig. 3](#); [Supplementary Information](#)). Hence, we refrained from explicitly inferring the timescale of sr3WO gains and losses because this inference would rely on the sr3 phylogram. Ignoring sr3WO, our

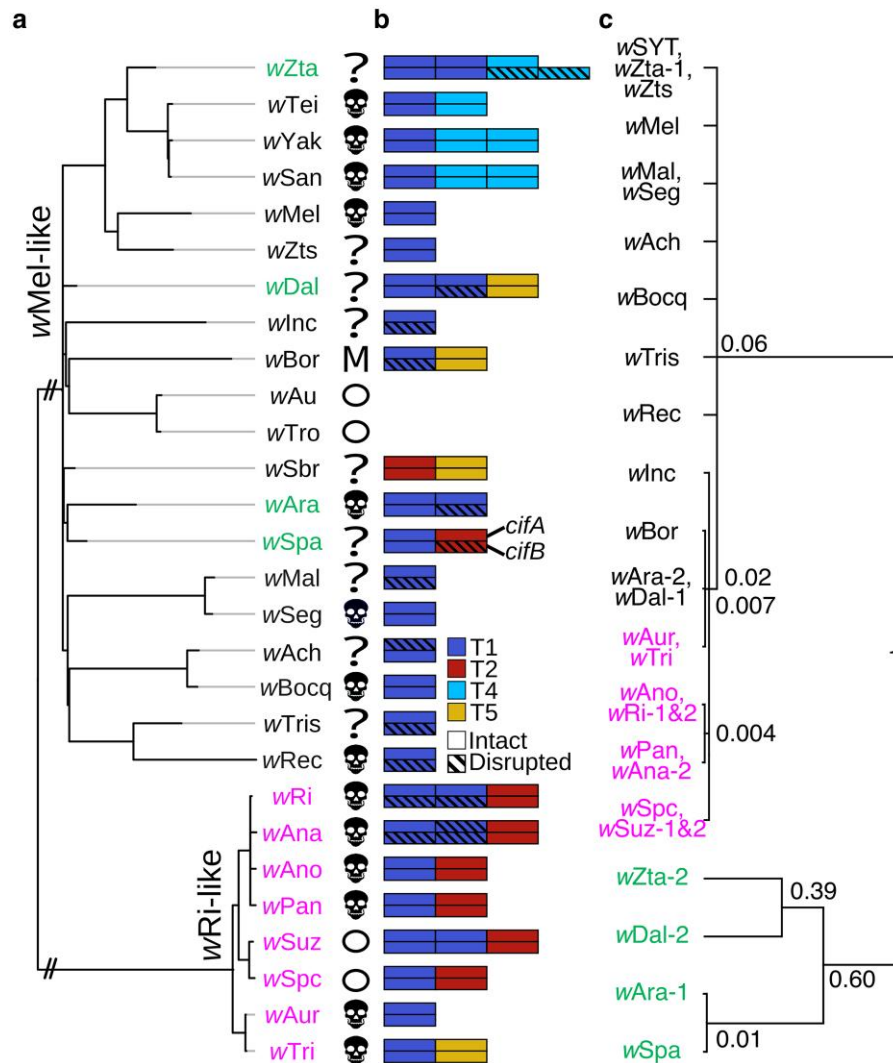


Fig. 5. Diverse *cif* operons rapidly turning over among *wMel*-like and *wRi*-like genomes. a) A phylogram of *wMel*-like (black and green) and *wRi*-like (pink) *Wolbachia*, including variants that cause CI (skull icon), do not cause CI (circles), or where CI status is unknown (?). *wBor* does not cause CI, but it does kill males (M) (Supplementary Table 8). Our new calibrations indicated that the *wRi*-like clade diversified over about 14 to 218 KY, the *wMel*-like clade diversified over about 206 KY to 2.4 MY, and the MRCA for the bifurcating genomic regions of *wMel* and *wRi* is about 1.4 to 22 MYA. Branches leading to these clades are shortened for presentation (//), and light gray branch extensions are used to improve visualization. b) Across *wMel*-like and *wRi*-like *Wolbachia*, 4 of the 5 characterized *cif* operon Types (T1 to T5) are observed (Supplementary Table 9). *cifA* (top) and *cifB* (bottom) schematics are presented with multiple copies adjacent to one another and ordered by their sequence integrity. The inset legend denotes *cif* Type by color, with integrity classified as intact (fully shaded) or disrupted (hashed) based on whether mutations split the gene into multiple ORFs or truncate the protein, causing loss of a functional domain. c) A relative chronogram for *cifA*_[T1] copies with node labels indicating relative ages, scaled to 1 for the root node. Identical sequences are collapsed into a single tip, and nodes with posterior probability < 0.95 are collapsed into polytomies. The majority of *cifA*_[T1] operons were closely related, with 4 divergent copies colored in green. Colors in panel c correspond to those for hosts in panel a. When multiple copies of the same Type were present in a genome and complete genomes were available, the dash and number following the variant name indicates the order in which copies occurred in their respective genomes. In cases where complete genomes are not available, copy numbers are assigned alphabetically based on NCBI accession names for contigs. All *cifA*_[T1] copies are represented, with the exception of *wAna*-1 that could not be confidently aligned. In the Supplementary Information we present phylograms for *cifA*_[T1] partitions based on GARD calls (Supplementary Fig. 4), which differed in the specific placement of the 4 outgroup copies.

conservative comparisons indicate that *Woviruses* based on sr typing can be gained and lost on the order of about 10^4 to 10^6 years.

Estimating *cif* turnover. CI is common (Fig. 5a), and *cif* complements are diverse (Fig. 5b) among *wMel*-like and *wRi*-like *Wolbachia*, with 14/19 characterized strains in our study causing CI (Fig. 5a; Supplementary Table 8; Supplementary Information) (Turelli et al. 2018; Martinez et al. 2021). This includes 2 *wMel*-like *Wolbachia* (*wSeg* in *D. seoyui* and *wBocq* in *D. bocqueti*), which we confirmed cause relatively intense CI (Supplementary Information). *wMel*-like and *wRi*-like *Wolbachia* carry between 1 and 3 *cif* operons that span 4 *cif* types (Fig. 5b; Supplementary Table 9; Supplementary Information). Applying our nematode-

based calibrations to comparisons of *cifs* observed in closely related *Wolbachia*, we confirmed that rapid *cif* turnover contributes to this diversity. Among *wSYT* variants, 2 *cif*_[T4] operons are observed in *wSan* and *wYak*, while *wTei* contains only one (Baiao et al. 2021). These 2 copies are identical, implying plausible duplication of a *cif*_[T4] operon after *wSYT* diverged, and prior to *wSan*-*wYak* divergence 10^3 to 10^4 years ago. *wMel* and *wZts* genomes did not contain *cif*_[T4] operons, indicating acquisition of *cif*_[T4] operons by *wSYT* occurred 96 KYA to 1.16 MYA. (With the exception of *wZta*, all genomes in this comparison are fully circularized.) In the ((*wAra*, *wSpa*), *wSbr*) clade (CPR: 134 KYA to 1.62 MYA), we observed turnover of multiple *cif* Types: *wAra*

Table 7. Estimates of root ages of 4 sets of *Wolbachia*: *wMel*-like ($N = 20$), *wRi*-like ($N = 8$), both clades combined ($N = 28$) (which span the 7 supergroup A *Wolbachia* described in our Materials and Methods and in our Results [(((*wMel*, *wAu*), (*wBic*, *wBar*)), ((*wRi*, *wTri*), *wHa*)))]), and A and B supergroup members from Fig. 1 in Meany et al. (2019) ($N = 15$ variants).

Comparison:	Plausible Time Estimate	
	Point	CPR
<i>wMel</i> -like root	889,363	(206,051; 2,433,090)
<i>wRi</i> -like root	76,421	(14,051; 218,089)
<i>wMel</i> -like and <i>wRi</i> -like root	7,545,595	(1,441,730; 22,004,200)
A to B supergroup root	151,570,339	(34,100,400; 414,809,000)

Plausible point estimates are calculated as the average of the point estimates of the fastest (*Dirofilaria*, lognormal) and slowest (*Brugia*, lognormal) nematode-based calibrations. We also report the CPR, which ranges from the lower-bound of the *Dirofilaria* range to the upper-bound of the *Brugia* range.

contained 2 *cif*_[T1] operons, its sister *wSpa* contained 1 *cif*_[T1] operon and 1 *cif*_[T2] operon, and outgroup *wSbr* contained only *cif*_[T2] and *cif*_[T5] operons. Similarly, in the (*wInc*, (*wBor*, (*wAu*, *wTro*))) clade (CPR: 126 KYA to 1.6 MYA), *wInc* and *wBor* each contained a *cif*_[T1] operon, but *wBor* contained a *cif*_[T5] operon, and *wAu* and *wTro* did not contain any *cif* operons. Finally, all *wRi*-like variants contained a single *cif*_[T2] operon, with the exception of *wAur* and *wTri*, that instead had a *cif*_[T5] operon absent from all other *wRi*-like genomes, including the fully circularized genomes of *wRi* and *wAna*. This implies turnover of both *cif* Types since the *wRi*-clade diverged 14 to 218 KYA (Table 7).

Because *cif*_[T1] operons are the most common *cif* Types in our dataset (Supplementary Table 9), we focused on homologs of *cifA*_[T1] to further illustrate rapid *cif* turnover. We present a relative chronogram for *cifA*_[T1] copies (Fig. 5c). We observed 2 distantly related clades of *cifA*_[T1] alleles (Fig. 5c). The first included 26 alleles observed across 16 *wMel*-like *Wolbachia* and 8 *wRi*-like *Wolbachia*. *wRi* and *wSuz* each carry 2 closely related *cifA*_[T1] alleles (Supplementary Table 10). In both *wRi* and *wSuz*, these alleles occurred 8,019 bp from an sr, implying ancestral duplication of the phage is plausible (Supplementary Information). *cifA*_[T1] alleles observed in *wMel*-like *wDal*, *wAra*, *wBor*, and *wInc* were most closely related to *cifA*_[T1] alleles observed in *wRi*-like *Wolbachia*. They shared particularly high identity with *cifA*_[T1] alleles observed in *wRi*-like *wAur* and *wTri* (99.2 to 99.8% aa identity). The second distantly related *cifA*_[T1] clade included additional *cifA*_[T1] copies observed in *wMel*-like *wZta*, *wDal*, and *wAra* genomes that were more closely related to each other than they are to the second *cifA*_[T1] copy in their respective *Wolbachia* genomes. The relationships of copies within this second clade, which also contained a single *cifA*_[T1] allele observed in *wSpa*, differed across 2 *cifA*_[T1] partitions based on GARD recombination calls (Supplementary Fig. 4; Supplementary Information). Finally, we observed that many *cifA*_[T1] copies of high sequence similarity spanned relatively deep phylogenetic divergences among plausibly associated sr3 alleles (Supplementary Fig. 5), a pattern indicative of *cif* turnover among Woviruses (LePage et al. 2017; Supplementary Information).

In summary, we conclude that *cif* gains and losses occurred in *wMel*-like and *wRi*-like genomes on the order of about 10^4 to 10^6 years, excluding cases that likely involved duplications, which occurred even more rapidly. In the Discussion, we consider plausible mechanisms that may have contributed to these patterns.

Estimating selection on *cifs*

CI phenotypes and the *cifs* that underlie them are common among *wMel*-like, *wRi*-like and other *Wolbachia* (Fig. 5a; Supplementary

Tables 8 & 9) (Shropshire et al. 2020a; Turelli et al. 2022). Putative loss-of-function mutations are more commonly observed and occur faster in *cifB* than in the *cifA* rescue locus (Martinez et al. 2021). These patterns also emerged from our comparisons of *Cif* protein similarity and identity (Supplementary Table 10). For example, *CifA*_[T1] proteins were more similar than *CifB*_[T1] proteins from the same pairs (Fig. 6a, Supplementary Figs. 6 & 7; Supplementary Information). To further understand these observations, we tested for selection during the divergence of putatively intact *cifA*_[T1] and *cifB*_[T1] alleles with no observed disruptions. Our analyses focus on type 1 loci because they are common among our *wMel*-like and *wRi*-like variants (Fig. 5b). For *cifB*_[T1], we also assessed selection on domains involved in CI and other functions, relative to its non-domain regions (Beckmann et al. 2017; LePage et al. 2017; Chen et al. 2019; Beckmann et al. 2021; Deehan et al. 2021; Kaur et al. 2022, 2024). These include 2 PD-(D/E)XK nuclease domains (Nuc1 and Nuc2) and a single deubiquitinase domain (Dub). We analyzed the estimated normalized ratio of non-synonymous to synonymous substitutions (ω), comparing *cifA*_[T1] and *cifB*_[T1], as well as *cifB*_[T1] domains to non-domain regions.

Across all pairwise contrasts, ω for *cifA*_[T1] (range: 0.26 to 0.38) was consistently and significantly lower than ω for *cifB*_[T1] (range: 0.34 to 0.46) (Fig. 6b; HL median difference = -0.06 ; Wilcoxon $V = 0$, $P = 0.02$; Supplementary Table 11), with Bayesian bootstrap posteriors indicating with high probability ($P = 1.0$) that *cifA*_[T1] ω was lower, i.e. $P(\Delta < 0)$ where $\Delta = \omega_{\text{cifA}} - \omega_{\text{cifB}}$. Sliding window analyses assessed in 3D space agreed, with pooled medians for ω across all windows lower for *CifA*_[T1] (0.29) than *CifB*_[T1] (0.37) ($\Delta_p = -0.23$; BCa 95% CI: $-0.38, -0.08$; $P < 0.0002$; Supplementary Information). Estimates of ω imply that selection acts to preserve intact copies of *cifA*_[T1] and *cifB*_[T1], with lower ω for *cifA*_[T1], implying that this locus involved in CI rescue is under greater constraint.

We evaluated 6 models in codeml that differ in their treatment of variation in ω across lineages and across sites. Prior to implementing these tree-based analyses, we tested for intralocus recombination using GARD, which found well-supported evidence for recombination in *cifA*_[T1] but not *cifB*_[T1] (Supplementary Fig. 4; Supplementary Information). We found no evidence for selection pressures differing across lineages for either gene or *cifA*_[T1] partitions (Supplementary Tables 12 & 13). The one-ratio model that assumes a single ω across all sites and lineages estimated an average $\omega = 0.39$ for *cifB*_[T1], while estimates for upstream and downstream *cifA*_[T1] partitions (P_1 and P_2 , respectively) differed from one another (P_1 average $\omega = 0.22$; P_2 average $\omega = 0.38$). This indicates that the constraint on *cifA*_[T1] is driven by relatively strong purifying selection on the N-terminus partition (sites 1 to 146). The N-terminus regions contribute to both CI and rescue (Shropshire et al. 2020b), while the C-terminus is important only for inducing CI (Shropshire et al. 2022).

We observed heterogeneity in ω across *cifB*_[T1] and *cifA*_[T1] partitions (Fig. 6d & e; Supplementary Tables 12 & 13; Supplementary Information), but models allowing a positively selected site class were supported only for *cifB*_[T1] and P_2 of *cifA*_[T1]. Among them, only *cifB*_[T1] sites 256, 258, and 341 were identified by both positive-selection models with Bayes Empirical Bayes (BEB) posterior $P \geq 0.95$ (Supplementary Table 14). Further analysis indicated that the site-specific *cifB*_[T1] results may be artifacts. Site 341 is located within the Nuc1 domain of *cifB*_[T1], while the other 2 sites are located in a 5' non-domain region. We evaluated alignments of all 3 sites for evidence of multi-nucleotide mutations (MNMs) that violate assumptions of the codon substitution model (Venkat

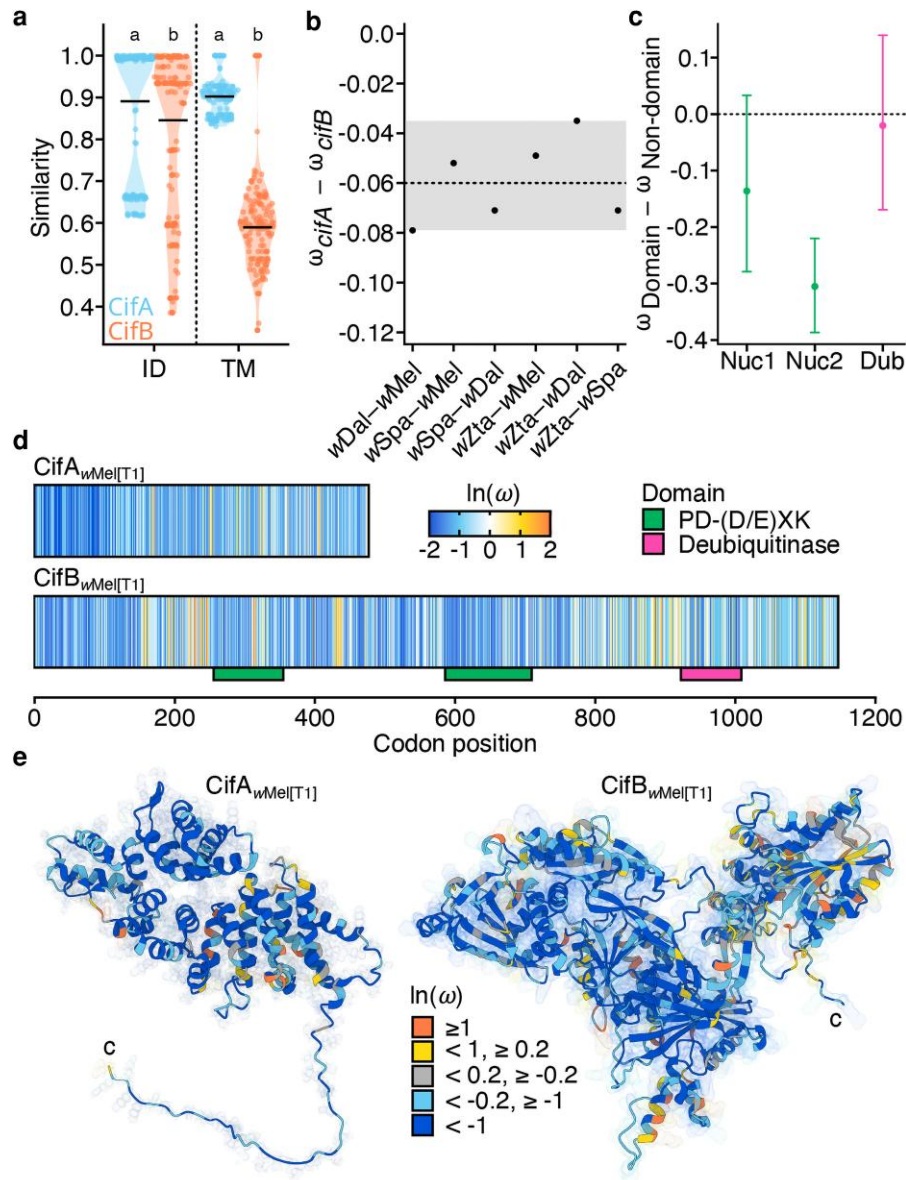


Fig. 6. Selection acts on intact *cifA*_[T1] and *cifB*_[T1]. a) Putatively intact *CifA*_[T1] proteins are more similar than intact *CifB*_[T1] from the same pairs ($N = 18$) in both sequence identity (ID) and structural similarity (TM). b) Pairwise median differences between *cifA*_[T1] and *cifB*_[T1] ω are consistently negative, with statistically significant $\omega_{cifA} < \omega_{cifB}$ across all contrasts ($P = 0.02$). The gray bar denotes the 95% CI for the Hodges-Lehmann estimator of the median of paired differences. c) Differences in mean per-site posterior ω between the 3 *cifB*_[T1] domains and pooled non-domain regions. The difference between one PD-(D/E)XK nuclease domain (Nuc2) and pooled *cifB*_[T1] non-domain regions was statistically significant ($P < 0.0001$). d) Linear schematic of the per-site posterior ω estimates across *CifA*_[T1] and *CifB*_[T1]. For *CifB*_[T1], colored boxes below the schematic denote HHPred-annotated functional domains as they occur along the codon alignment. e) 3D schematic of the per-site posterior ω estimates mapped onto *CifA*_[T1] AlphaFold2 structures. "C" denotes the C-terminus of each protein. Per-site posterior ω estimates were obtained from the best-supported model for each gene and partition based on likelihood-ratio tests (Supplementary Tables 12 & 13). As noted in our text, all sites identified by positive-selection models with BEB posterior $P \geq 0.95$ show evidence of MNMs. Colors indicate $\ln(\omega)$ values such that neutrality ($\omega = 1$) maps to 0: dark blue (< -1), light blue ($-1 \leq x < -0.2$), gray ($-0.2 \leq x \leq 0.2$), yellow ($0.2 < x \leq 1$), and orange (> 1).

et al. 2018). We observed 4 unique codons at position 341: tgg (Trp) in wMel, ttt (Phe) in wDal, agg (Arg) in wSpa, and gga (Gly) in wZta. All pairwise contrasts involve at least 2 nucleotide changes, with the exception of tgg $\leftrightarrow$ agg. All other BEB-supported sites have at least 2 contrasts that appear MNM-like.

Given that all analyses supported preservation of intact *cifB*_[T1], we summarized codon ω estimates by functional domain and compared them to the pooled non-domain background using label-permutation tests. Using the site-level posterior estimates for ω from the beta plus positive-selection model, we observed mean posterior $\omega < 1$ for Nuc1 (0.55, 95% bootstrap interval: 0.42 to 0.71), Nuc2 (0.38, 95% bootstrap interval: 0.33 to 0.45), Dub

(0.67, 95% bootstrap interval: 0.53 to 0.81), and for non-domain regions (0.69, 95% bootstrap interval: 0.63 to 0.75). Only the mean posterior difference involving Nuc2 was statistically significant ($\Delta = -0.31$; 95% bootstrap interval: $-0.39, -0.22$; $P < 0.0001$; Fig. 6c). The observed difference for Nuc1 was also negative ($\Delta = -0.14$), but with a relatively wide bootstrap interval that included 0 (95% bootstrap interval: $-0.28, 0.03$; $P = 0.06$). In contrast, the mean posterior difference between the Dub and non-domain regions was small and statistically nonsignificant ($\Delta = -0.02$; 95% bootstrap interval: $-0.17, 0.14$; $P = 0.43$), reflecting their similar mean posterior estimates. Sliding-window analyses in 3D space broadly agreed, with the mean of the per-pair median differences

on the log scale negative for all comparisons, but only statistically significant for Nuc2 ($\Delta_p = -0.38$; $P = 0.003$, [Supplementary Information](#)).

In summary, all analyses supported *cifA*_[T1] being more constrained than *cifB*_[T1], with constraint on *cifA*_[T1] driven by relatively strong purifying selection on the N-terminus regions involved in CI rescue ([Shropshire et al. 2020b](#)). However, our analyses indicated that all regions of intact *cifB*_[T1] are also constrained, with particularly strong purifying selection on the Nuc2 domain. We discuss these results in light of several theoretical predictions and the timescale of *cifB* degradation ([Beckmann et al. 2021](#)) and the loss of CI ([Meany et al. 2019](#); [Martinez et al. 2021](#)).

Discussion

Decades of progress have increased our understanding of the timescale and processes of *Wolbachia* acquisition and the nature of *Wolbachia* genomic evolution. However, time estimates have remained uncertain because there have been few well-established examples of codiverging *Wolbachia* and hosts, combined with fossil-based estimates of host divergence times. Here, we used recently available chronograms for filarial nematodes and complete sequences of their codiverging, obligately symbiotic *Wolbachia* to provide new calibrations and conservative estimates of their uncertainty. We also provide detailed analyses of *Wolbachia* genomic evolution, focusing separately on *cifs* that cause and rescue CI, the *Wovirus* prophages that contain *cifs*, and single-copy, non-phage *Wolbachia* genes. Our results demonstrate that these are temporally nested sets, with *Wolbachia* moving rapidly among hosts and *Wovirus* and *cifs* moving even more rapidly among *Wolbachia* genomes. Along their journey, efficient selection maintains *cifA*, while weaker stabilizing selection maintains intact *cifB*. Although the clade boundaries for the young *wMel*-like and *wRi*-like *Wolbachia* are arbitrary, and the range of plausible calibrations is wide, our central conclusions are robust to uncertainty in timing and incomplete sampling of these clades. We discuss our major findings.

Timing the mosaic and bifurcating evolution of some *Wolbachia* genomes

Ever since the discovery of intragenic recombination involving genes from distantly related *Wolbachia* ([Jiggins et al. 2001](#); [Werren and Bartos 2001](#); [Baldo et al. 2006](#); [Ellegaard et al. 2013](#); [Wang et al. 2020](#)) and the availability of complete *Wolbachia* genomes (including *wMel*, [Wu et al. 2004](#); *wPip*, [Klasson et al. 2008](#); and *wRi*, [Klasson et al. 2009](#)), *Wolbachia* have been understood as “mosaics” because different portions of their genomes (and different portions of individual genes) follow discordant phylogenies. We here estimate the timing and genomic extent of that mosaicism, applying new divergence-time estimates from 2 distinct calibrations to 4 sets of *Wolbachia*: some closely related to *wRi* (with conservative plausible range [CPR] for their divergence times: 14 to 218 KYA), some closely related to *wMel* (CPR: 0.2 to 2.4 MYA), 7 supergroup A *Wolbachia* spanning the divergence of *wMel* and *wRi* (CPR: 1.4 to 22 MYA), and the highly diverged *Wolbachia* from filarial nematodes that provide our calibrations ([Table 7](#)).

Within each of these clades, including the youngest, the *wRi*-like variants, we document extensive movement of phages and *cifs* that occur within those phages. Hence, a significant fraction of these *Wolbachia* genomes is clearly “mosaic.” However, when we consider only single-copy, non-phage loci, we find no evidence of horizontal acquisition of distantly related alleles within the *wRi*-like variants. Increasing divergence time by an order of magnitude to our *wMel*-like clade, we also observe no plausible

evidence for horizontal allele acquisition. Zooming out another order of magnitude to 7 supergroup A *Wolbachia* that span the divergence between *wMel* and *wRi*, we find a single example of horizontal acquisition involving an allele homologous to RefSeq_WP_010962975.1 in *wHa*. While the *wHa* allele differs substantially from the homologous alleles in the other 6 supergroup A *Wolbachia* (18.5% to 18.9%), it is identical to copies in *Wolbachia* from several diverse host species. Given that the divergence time between the *wMel*-like and *wRi*-like clades is more than twice as old as the crown age of either clade, our 2-fold outlier criterion would have identified exchanges between them.

We did find some evidence of phylogenetic discordance from individual single-copy genes among these 7 genomes spanning the divergence of *wMel* and *wRi*, but there is still a strong signal for a single topology. The sources of discordance could be ILS-like effects associated with short branches deep in the tree or model mis-specification leading to spurious inferences. Our data suggest that horizontal acquisition of distantly related alleles involves very few single-copy loci over 1 to 2 MYA. We conclude that *Wolbachia* genomes contain bifurcating genomic regions composed of single-copy loci, from which we can robustly estimate divergence times. Moreover, we find no evidence of recombination or horizontal acquisition of distantly related alleles for the supergroup C and D *Wolbachia* from filarial nematodes on which we base our calibrations.

Our results do not contradict previous evidence of extensive recombination involving relatively distantly related *Wolbachia*—because of the very different timescales involved. However, they do contradict the claim of [Ellegaard et al. \(2013\)](#) that extensive recombination among genomes within supergroups A and B produces highly discordant phylogenies from single-copy loci. Using only 3 variants within each supergroup and low stringency for node resolution (>75% bootstrap support) in their single-gene phylogenetic analyses, [Ellegaard et al. \(2013\)](#) found nearly equal support for all 3 alternative rootings within each supergroup ([Fig. 3](#)). Under purely bifurcating evolution, 3 taxa possess only 1 unrooted topology. Their rooting for the 3 variants within each supergroup is based on comparisons to distantly related outgroups from the other supergroup. Given the deep A to B split ([Table 7](#)), this rooting is unreliable because of saturation artifacts related to long-branch attraction ([Brinkmann et al. 2005](#)). Empirical and simulation studies show that these “distant outgroup” artifacts can be resolved by analyzing more ingroup taxa and using more closely related outgroups (cf. [Rannala et al. 1998](#); [Heath et al. 2008](#), p. 245). Our analyses do both. They indicate that the [Ellegaard et al. \(2013\)](#) results may reflect their inability to resolve the root position for each supergroup, rather than pervasive phylogenetic discordance involving single-copy genes within each supergroup.

Overall, we find little evidence of “mosaicism,” or phylogenetic discordance, among the single-copy genes we examine within and between our relatively recently diverged *wMel*-like and *wRi*-like clusters. This contrast, between extensive evidence for recombination within single-copy loci between distantly related *Wolbachia* (e.g. the [Klasson et al. \(2009\)](#) comparison of *wRi* to distantly related supergroup A and B *Wolbachia*), and little evidence for such recombination between closely related variants, parallels the finding that although horizontal *Wolbachia* acquisition is clearly pervasive among distantly related hosts, it seems rare among individuals within specific host species ([Richardson et al. 2012](#); [Cooper et al. 2019](#)), with some possible exceptions ([Scholz et al. 2020](#)). Rare horizontal transmission of *Wolbachia* within host species does not affect our divergence-time estimates of *Wolbachia*

variants found in distinct host species—nor do rare examples of horizontal acquisition of single-copy *Wolbachia* loci. Scholz et al. (2020) noted extreme variation in relative rates of mtDNA vs *Wolbachia* divergence, but much greater consistency between nuclear and *Wolbachia* divergence. Our nematode calibrations, which show only 3-fold rate variation among distantly related *Wolbachia*, are consistent with this observation.

Estimating *Wolbachia* divergence times

Our nematode-based *Wolbachia* rate calibrations, combined with new and existing genomic data, enabled us to revisit hypotheses concerning *Wolbachia* acquisition by *Nasonia* wasps and *Nomada* bees and to quantify the timescale of *Wolbachia* movements among distantly related hosts. Our analysis of the relatively young *wMel*-like and *wRi*-like *Wolbachia* clades (Figs. 1a, 4a, 5a & Supplementary Fig. 8), focused on the first *Wolbachia* identified in *Drosophila* species (Hoffmann et al. 1986; Hoffmann 1988), suggest that many young *Wolbachia* clades (i.e. ones that diverged on the order of 1 MYA or less) spread rapidly among anciently diverged hosts (that diverged on the order of 50 to 350 MYA). The host range and rapid movement of *wMel*-like variants parallel those of *wRi*-like variants (Turelli et al. 2018), suggesting that this scenario—in which closely related *Wolbachia* invade anciently diverged hosts—may characterize many common *Wolbachia*. Both CI phenotypes and *cif* complements are diverse among relatively recently diverged *Wolbachia*, with *Woviruses* and their associated *cifs* gained and lost on the order of 10^4 to 10^6 years.

An obvious question is whether our calibrations based on obligate C and D supergroup *Wolbachia* in nematodes are applicable to the facultative A and B supergroup *Wolbachia* in insects. The strongest support for this extrapolation comes from the fact that the range of point estimates for third-site substitution rates from nematodes, 6.47×10^{-10} to 1.82×10^{-9} , closely match the most plausible point estimates based on *Nomada* and *Nasonia*, namely 6.46×10^{-10} and 2.2×10^{-9} , respectively. The *Nasonia* estimate rests on separate molecular clock estimates from insects and bacteria rather than fossil calibration. Obtaining more reliable point estimates and levels of uncertainty for *Wolbachia* substitution rates requires additional examples of *Wolbachia*-host codivergence, that is, instances in which specific *Wolbachia* infections persist through host speciation. Such examples will come from increasingly reliable chronograms for insects, based on dated, amber-preserved fossils, and genome-based estimates of host and *Wolbachia* phylogenies and levels of sequence divergence, including mtDNA data. Our new filarial-nematode-based calibrations contribute to this goal. The rates estimated are significantly slower than the mtDNA-mutation-based estimate proposed by Richardson et al. (2012). Our chronograms are based on these new calibrations, but Table 5 compares our new time estimates to the shorter ones based on Richardson et al. (2012).

Three points are worth noting about these alternative calibrations. First, for eukaryotic nuclear genomes, some data indicate that substitution rates per year for synonymous substitutions and introns can closely match yearly mutation rate estimates. See, for instance, Table 1 of Begun et al. (2007) for sequence divergence along the terminal branches leading to *D. melanogaster* and *D. simulans*. The rates per year per site are quite comparable to the *D. melanogaster* mutation rate estimates of Keightley et al. (2009), assuming roughly 10 generations per year and that *D. melanogaster* and *D. simulans* diverged about 4 MYA.) Second, as noted by Ho et al. (2005), mtDNA substitution rates slow dramatically over a timescale of tens of thousands of years. Finally, even if the faster Richardson et al. (2012) calibration is more accurate for the

wMel-like and *wRi*-like variants we analyzed than our new calibrations, our key conclusions remain valid: namely, many closely related *Wolbachia* move rapidly between distantly related hosts; and over that rapid between-host-movement timescale, *Woviruses* and *cifs* within those *Wolbachia* turn over even more rapidly.

It is worth revisiting the Meany et al. (2019) estimate of supergroup A and B divergence with our new calibration. Using the Richardson et al. (2012) rate and both relaxed-clock [branch-rate priors $\Gamma(2,2)$ and $\Gamma(7,7)$] and strict-clock analyses, Meany et al. (2019) estimated a plausible divergence interval for supergroups A and B of 6 to 64 MY. This range overlaps that of Werren et al. (1995), based on applying a “universal” bacterial synonymous substitution rate (Ochman and Wilson 1987) to the *ftsZ* locus (59 to 67 MYA). Both the Meany et al. (2019) and Werren et al. (1995) A to B divergence-time estimates were more recent than the Gerth and Bleidorn (2016) estimate of 76 to 460 MYA (point estimate: 217 MYA). However, both the Meany et al. (2019) and Werren et al. (1995) estimates now seem too recent, because the Richardson et al. (2012) calibration seems too fast and the *ftsZ* locus evolves slowly relative to most bacterial genes (Supplementary Information). Meany et al. (2019) questioned the Gerth and Bleidorn (2016) estimate for two reasons. First, their chronogram placed *wNo* (from *D. simulans*) sister to all other supergroup B *Wolbachia*, which disagreed with their own phylogram and others (Lindsey et al. 2018). Second, their estimate of *wRi*-*wSuz* divergence of 0.9 MY (with support interval 0.2 MY to 2.2 MY) was nearly 2 orders of magnitude higher than the 11 KY estimate of Turelli et al. (2018). However, Turelli et al. (2018) used the Richardson et al. (2012) rate that we think underestimated *wRi*-like *Wolbachia* divergence by about a factor of 7. Using our new calibrations, we now estimate *wRi*-*wSuz* divergence as 76 KYA (CPR: 14 to 218 KYA) and supergroup A and B divergences as 152 MYA (CPR: 34 to 415 MYA) (Table 7). Our wide plausible ranges overlap those of Gerth and Bleidorn (2016); but as described below, we believe their point estimates are too large because their calibrations conflate cladogenic and horizontal *Wolbachia* acquisition.

Our analyses focus on the timescale over which clades of closely related *Wolbachia* have been acquired by distantly related hosts. In contrast, Bailly-Bechet et al. (2017) used a sample of about 1,000 arthropod species (specifically OTUs characterized by mtDNA CO1 sequences) from Pacific islands, to estimate how long, on average, specific *Wolbachia* infections persist in host lineages and how long uninfected host lineages remain uninfected. Bailly-Bechet et al. (2017) inferred that *Wolbachia* tend to persist in lineages for approximately 7 million years. However, as Cooper et al. (2019) noted, this estimate is inconsistent with the fact that few sister species share codiverging *Wolbachia*. Because speciation times for *Drosophila*, and presumably many other insect clades, are on the order of 1 million years or less (Coyne and Orr 1997; Turelli et al. 2014), the infection-duration estimated by Bailly-Bechet et al. (2017) would lead to codivergence of *Wolbachia* and hosts for many sister pairs.

The estimated *Wolbachia* substitution rates, with conservative plausible ranges, from *Nomada* (Table 4) suggest a history of both cladogenic and horizontal *Wolbachia* acquisition in this 4-species clade. As shown in Table 4, the *Wolbachia* substitution rate based on comparisons of *N. ferruginata* and the 3-ingroup species agrees with that obtained from the *Brugia* nematode pair. In contrast, assuming cladogenic inheritance of *Wolbachia* throughout the 4-species *Nomada* clade implies a greater than 10-fold slowdown of *Wolbachia* divergence between *N. flava* and *N.*

leucophthalma (which diverged about 0.8 MYA, see Table 4) vs the outgroup *N. ferruginata* (from which they diverged about 2.1 MYA), while their nuclear and mtDNA continued to diverge at a rate typical for the 4-species clade as a whole (see Table 3). Such a dramatic recent slowdown of *Wolbachia* divergence seems implausible, given the relative constancy over a much longer timescale in the nematodes. We suggest that horizontal transmission of *Wolbachia* across the 3-species ingroup is more plausible.

As noted in our discussion of Fig. 4b, individual genomes do not allow us to determine the direction or temporal order of *Wolbachia* acquisitions among hosts. Resolution is possible when multiple *Wolbachia* genomes are available from each host. If the *Wolbachia* haplotypes in species A are nested within the phylogeny of *Wolbachia* in species B, we can reasonably infer that species A acquired *Wolbachia* from species B. Conversely, if the closely related *Wolbachia* and mtDNA in host species A and B are reciprocally monophyletic sister clades with comparable divergence times, we can be reasonably confident of cladogenic *Wolbachia* acquisition. However, given the long time required to achieve reciprocal monophyletic gene trees—at least under neutrality (Hudson and Coyne 2002)—cladogenic *Wolbachia* transmission may often produce tangled trees in which at least 1 of the host species carries *Wolbachia* (and mtDNA) haplotypes that are paraphyletic with respect to those in a closely related host. The relative frequencies of these scenarios can be determined only with additional genomic analyses of the *Wolbachia*, mtDNA and nuclear genomes of closely related hosts.

Turelli et al. (2018) analyzed multiple *Wolbachia* sequences from several species, documenting rapid horizontal transfer of wRi-like *Wolbachia* between divergent *Drosophila*. Specifically, the wRi sequences from *D. simulans* formed a clade nested within a paraphyletic cluster of wAna sequences from *D. ananassae*. Given that *D. ananassae* and *D. simulans* span the *D. melanogaster* species group, which diverged on the order of 47 MYA (Fig. 1b), introgression is impossible. Hence we conjectured that horizontal non-sexual transmission of *Wolbachia* occurred between *D. ananassae* and *D. simulans*, with *D. ananassae* as a plausible source. Turelli et al. (2018, Fig. 1) also analyzed 8 *Wolbachia* sequences from *D. sukukii* (wSuz) and a single sequence from *D. subpulchrella* (wSpc). Two copies of wSuz from Asia (the other wSuz samples came from North and South America) formed a clade with the single wSpc, sister to the other 6 wSuz sequences. We used these data to conjecture that *D. sukukii* was a source of the *Wolbachia* in *D. subpulchrella*. However, the 2 putative Asian wSuz sequences, obtained from GenBank, were in fact derived from *D. subpulchrella* specimens (China-AA27 and Korea-AA7). This misidentification was first kindly pointed out to us by Mathieu Gaultier, then confirmed by Joanna Chiu. With this correction, the data in Fig. 1 of Turelli et al. (2018) produce sister clades of 6 wSuz and 3 wSpc, providing no information on the direction of transfer (Supplementary Information).

Estimating the timescale of Wovirus and cif movements

The rapidly host-switching *Wolbachia* in our study differ widely in their CI phenotypes, from variants that do not cause CI to others that cause relatively weak or intense CI (Fig. 5a; Supplementary Table 8). This includes wMel-like wSeg and wBocq that we confirmed cause relatively intense CI (Supplementary Information). CI can accelerate *Wolbachia* spread within host species (Hoffmann et al. 1990; Hoffmann and Cooper 2024), as observed for wRi that rapidly spread through global *D. simulans* populations to relatively high and stable frequencies (Hoffmann et al. 1986,

1990; Turelli and Hoffmann 1991, 1995; Kriesner et al. 2013). CI may also increase the likelihood of interspecific *Wolbachia* transfers, like those we have observed, through a process of clade selection (Hurst and McVean 1996; Turelli et al. 2022). Unsurprisingly, the effectors of CI have been of intense interest since the discovery that *Wolbachia* cause CI in *C. pipiens* (Yen and Barr 1971, 1973). The relatively recent discovery of cifs (Beckmann et al. 2017; LePage et al. 2017; Shropshire et al. 2018) facilitated genomic analyses that have recently characterized cif diversity (Lindsey et al. 2018; Turelli et al. 2018; Martinez et al. 2021; Vancaester and Blaxter 2023; Amoros et al. 2025a). It is well established that Wovirus and cif diversity reflect extensive horizontal transfers within and among *Wolbachia* lineages (e.g. Masui et al. 2000; Bordenstein and Wernegreen 2004; Kent et al. 2011a, 2011b; Ellegaard et al. 2013; LePage et al. 2017; Cooper et al. 2019; Madhav et al. 2020; Martinez et al. 2021; Bordenstein and Bordenstein 2022; Vancaester and Blaxter 2023; Amoros et al. 2025a). What has remained unresolved is the timescale over which these exchanges occur. We estimate that Wovirus and cif exchanges involving wMel-like and wRi-like *Wolbachia* occur on the order of 10^4 to 10^6 years, excluding more rapid changes plausibly involving duplications. These bounds are conservative and derive from our nematode-based divergence estimates of bifurcating genomic regions and comparisons of cifs observed in closely related *Wolbachia*. High-quality Wovirus assemblies from additional *Wolbachia* genomes are required to further resolve the timing of these gains and losses.

The most plausible mechanisms of rapid cif exchanges include natural transformation and movements involving Wovirus (Bordenstein and Wernegreen 2004; Gao et al. 2022; Vancaester and Blaxter 2023), plasmids (Martinez et al. 2022), and transposons (Gillespie et al. 2018; Cooper et al. 2019). While characterization of *Wolbachia* plasmids remains limited (Reveillaud et al. 2019), IS elements are abundant and transcriptionally active in *Wolbachia* genomes (Duron et al. 2005; Iturbe-Ormaetxe et al. 2005; Riegler et al. 2005; Cordaux et al. 2008; Klasson et al. 2009; Ellegaard et al. 2013; Cerveau et al. 2015), and frequently occur near cifs in Wovirus (Supplementary Tables 15–17). IS-mediated transfer was first identified as a plausible mechanism involved in the acquisition of cif_[T4] loci by Wovirus observed in wYak and closely related genomes (Cooper et al. 2019) (Fig. 5a; Supplementary Information). Evidence for IS involvement in cif transfer has subsequently been documented in other *Wolbachia* (Madhav et al. 2020) and in diverse *Orientia tsutsugamushi* (Oswalt et al. 2025). These exchanges resemble cases where IS element-based composite and other compound transposons (Berg and Howe 1989) are responsible for moving antibiotic resistance genes among bacterial species (Sigquier et al. 2014; Razavi et al. 2020). cif exchanges mediated by IS elements most likely occur when *Wolbachia* co-occur and actively replicate in the germline, enabling IS elements to mobilize cif modules and generate short extrachromosomal DNA fragments, which can subsequently integrate via recombination or other end-joining pathways into the recipient. Other transposons have been implicated in horizontal cif transfers, including PDDEXK2-family transposases that occur near cif loci (Supplementary Table 15) in predicted transposon-like modules (Tan et al. 2024; Amoros et al. 2025a). These transposases cluster phylogenetically with cifs that span 8 bacterial genera and show a history of horizontal transfer (Amoros et al. 2025b). Hence, transposons along with well-documented contributions of Wovirus transfers and recombination, and potentially other mechanisms, contribute to the rapid cif turnover we document. Determining the timescale of cif

exchanges involving other symbionts (Gillespie et al. 2018; Amoros et al. 2025b) will require estimating their divergence times using reliable calibrations.

Estimating natural selection on *cifs* and assessing the maintenance of CI

Our analyses of selection on *cif* loci are based on comparing synonymous vs non-synonymous substitutions for intact *cif*_[T1] pairs across *Wolbachia* found in different hosts. We find that intact *cifA*_[T1] is more constrained than intact *cifB*_[T1] (Fig. 6; Supplementary Information). Our analyses support relatively stronger purifying selection on the 5' region of *cifA*_[T1] (Fig. 6d & e). Mutagenesis of CifA proteins previously found that N-terminus regions contribute to both CI induction and rescue (Shropshire et al. 2020b), while the C-terminus is important only for inducing CI. This suggests that selection to maintain functional *cifA*_[T1] is related to CI rescue. We also find that weaker stabilizing selection maintains *cifB*_[T1] involved in producing CI (Fig. 6d & e). Constraint on intact *cifB*_[T1] is driven in part by relatively strong preservation of its Nuc2 domain relative to pooled non-domain regions (Fig. 6c). In contrast, we estimate similar ω for Nuc1, Dub, and pooled non-domain regions that are each also subject to weaker stabilizing selection. Overall, we estimate stronger selection to maintain CI rescue than to maintain CI induction.

How do these empirical patterns of selection compare to theoretical predictions? It is critical to recognize that different evolutionary predictions concerning CI induction and CI rescue emerge from analyses of selection among *Wolbachia* variants within a specific host species vs “clade selection” analyses that consider the differential proliferation of *Wolbachia* variants across host species. Theory predicts that for *Wolbachia* lineages within an individual host species, natural selection acts to maintain resistance to CI, but it does not act to maintain the ability to induce CI against uninfected hosts (Prout 1994; Turelli 1994). In contrast, there is direct selection for *Wolbachia* to enhance host fitness (Prout 1994; Turelli 1994). These theoretical predictions are robust to population subdivision within a host species (Haygood and Turelli 2009). Even with extreme population subdivision, *Wolbachia* variants that greatly increase the intensity of CI will be selected against if they induce even very small costs on host fitness. In contrast, when considering which *Wolbachia* variants are most likely to spread among host species, an inherently slower “clade selection” process (Lewontin 1970), those that induce CI are favored. For *Wolbachia* that initially decrease host fitness, this is explained in terms relative chances of establishment in new hosts (Hurst and McVean 1996). In contrast, for *Wolbachia* that enter new hosts as mutualists, CI is favored because it produces higher infection frequencies and persistence times within host species (Turelli et al. 2022). Selection within host species always acts to maintain CI rescue, whereas selection to induce CI is driven primarily by weaker “clade selection.” These predictions are broadly consistent with evidence for stronger selective constraints on *cifA*_[T1] than intact *cifB*_[T1].

The contributions of specific *cifB*_[T1] domains to CI induction are unresolved. The CifB Dub includes the only known catalytic Cif motif, and mutating its conserved cysteine active site ablates CI (Beckmann et al. 2017), although it does not consistently ablate yeast toxicity (Terretaz et al. 2023). Naturally occurring single-nucleotide variants in the Dub can also alter CI intensity, as observed for a valine-to-leucine variant in *wYak cifB*_[T1]. This single carbon change reduces in vitro deubiquitylating efficiency and also the intensity of CI when expressed transgenically in *D. melanogaster* (Beckmann et al. 2021), potentially underlying observed

differences in *wYak* and *wMel* CI intensities (Hoffmann 1988; Cooper et al. 2017; Beckmann et al. 2021). While the CifB_[T1] Nuc1 and Nuc2 PDDEXK homologs lack canonical PD-(D/E)XK catalytic residues (Beckmann et al. 2017; Oladipupo and Hochstrasser 2025), mutating conserved sites in the *wMel* copy ablated CI (Shropshire et al. 2020a, 2020b; Kaur et al. 2024), and this superfamily exhibits substantial functional diversity, with some members retaining nuclease activity despite non-canonical motifs (Knizewski et al. 2007). Our results support that these *cifB*_[T1] domains (and non-domain regions) are preserved in intact *cifB*_[T1] copies, with particularly strong constraint on Nuc2. Despite this preservation of intact *cifB*_[T1] copies, putative loss-of-function mutations are known to be more common and to occur faster in *cifB* relative to *cifA* (Martinez et al. 2021). This includes observations of the loss of CI by *wMau* in *D. mauritiana* via *cifB* degradation and putative preservation of *cifA* (Meany et al. 2019).

Clade selection, as discussed above, can explain our observation of selective constraint on *cifB*_[T1]. Our analysis is restricted to putatively intact *cifB*_[T1] copies, and most loss-of-function mutations in *cifB* have occurred relatively recently, tending to occur on terminal gene-tree branches (Martinez et al. 2021). It is also clear that CifB contributes to other functions besides CI induction, and such pleiotropy could contribute to the selective constraint we observe (Turelli 1994). For example, Cifs are known to interact with host chromatin (Terretaz et al. 2023), and CifB seems to modulate *wMel* interactions with host autophagy to maintain *wMel* density in certain tissues (Deehan et al. 2021; Strunov et al. 2022). Given that *Wolbachia* densities in host tissues influence diverse host phenotypes and *Wolbachia* transmission rates (Bordenstein et al. 2006; Martinez et al. 2014; Chrostek and Teixeira 2015; Hague et al. 2022, 2024; Porter and Sullivan 2023; Radousky et al. 2023), modulation of host autophagy could significantly affect host and *Wolbachia* fitness. Regardless of whether the CifB Nuc2 domain functions enzymatically or through chromatin binding, the selection we observe is consistent with significant functional constraint. Those functions may involve CI, pleiotropic effects (e.g. involving host autophagy), or both. Future work focused on pleiotropic effects of CifB and its domains will contribute to our understanding of the preservation of intact *cifB*_[T1], which nevertheless seems destined to degrade through evolutionary time faster than *cifA*_[T1]. Similarly, phylogenetic analyses of the persistence of the CI phenotype within *Wolbachia* lineages, such as our *wRi*-like and *wMel*-like clades, can quantify the relative persistence times of *Wolbachia* lineages that do or do not cause CI.

Conclusions and implications for *Wolbachia* biocontrol

While the specific ecological mechanisms underlying rapid *Wolbachia* host switching in these closely related clades are unresolved, species interactions—and especially those involving parasitoid wasps (Vavre et al. 1999; Huigens et al. 2004; Gehringer and Vorburger 2012; Ahmed et al. 2015) and plant hosts (Caspi-Fluger et al. 2012; Chrostek et al. 2017; Li et al. 2017; Shi et al. 2024)—are plausible mechanisms underlying horizontal endosymbiont transfer (Cordaux et al. 2001; Kittayapong et al. 2003; Sintupachee et al. 2006; Jaenike et al. 2007; Oliver et al. 2010; Hurst et al. 2012; Hoffmann and Cooper 2024). Regardless of the mechanisms, pervasive and rapid *Wolbachia* host switching helps explain why *Wolbachia* are the most common invertebrate endosymbionts known and provides the physical opportunity for genomic exchanges between *Wolbachia*. These genomic exchanges also occur rapidly, with unique *cifs* sometimes observed in *Wolbachia* that diverged as recently as 10³ to 10⁴ years, with

evidence mounting for transposon involvement in these rapid transfers. These movements produce unique *cif* complements among closely related *wMel*-like and *wRi*-like variants and other *Wolbachia* that contribute to observed variation in CI intensity. Because CI directly contributes to *Wolbachia* spread within host populations (Hoffmann et al. 1990; Turelli and Hoffmann 1991), and may contribute to the likelihood of *Wolbachia* movements between host species (Hurst and McVean 1996; Turelli et al. 2022), understanding the phylogenetic distributions of CI and the loci that underlie it remains critical.

CI enables *Wolbachia*-based suppression of vector and pest populations and the alteration of natural vector populations with pathogen-blocking *Wolbachia* transinfections. When *wMel* *Wolbachia* are moved from *D. melanogaster* to *Aedes aegypti*—species that diverged about 250 MYA—the transinfection induces strong CI that can drive the endosymbiont to high frequencies (Hoffmann et al. 2011; Walker et al. 2011). *wMel* introductions into *Ae. aegypti* have been very effective in locations where they have spread to high, stable frequencies (Hoffmann et al. 2011; Utarini et al. 2021; Lenharo 2023; Velez et al. 2023), with one study reporting an 86% reduction in dengue in an Indonesian city (Utarini et al. 2021) and a reduction of greater than 90% in 2 cities (Lenharo 2023; Velez et al. 2023). However, *wMel* has been lost in some release locations (Hien et al. 2021; Moledo Gesto et al. 2021), particularly under very hot conditions. Temperature can affect CI intensity (Reynolds and Hoffmann 2002; Ross et al. 2019; Shropshire et al. 2021b) and the fidelity of maternal *Wolbachia* transmission (Ross et al. 2017; Hague et al. 2022, 2024). Identifying strong-CI-causing and virus-blocking *Wolbachia* from the tropics may facilitate *Wolbachia* biocontrol, and *wMel*-like variants that naturally associate with tropical host species are obvious candidates for new applications (Gu et al. 2022; Gong et al. 2023; Thia et al. 2025). Notably, *wMel*-like *Wolbachia* with diverse *cif* profiles have naturally colonized ecologically and geographically diverse hosts that are approximately 100 MY more diverged than are donor *D. melanogaster* and recipient *Ae. aegypti* dipterans (MRCA about 250 MYA, Wiegmann et al. 2011). We report a tropical *Wolbachia* variant (*wZts*) that, to our knowledge, is the closest relative to *wMel* in *D. melanogaster* (Supplementary Information). While *wZts* CI has not yet been tested, all 6 *Z. tsacasi* sampled in nature carry *wZts* (producing 0.61 as the 95% lower bound for *wZts* frequency), consistent with strong CI. Identification of *wZts* and other variants closely related to pathogen-blocking *wMel* has the potential to add versatility and contributes towards the development of more customized and environment-specific *Wolbachia* applications (Hoffmann and Cooper 2025).

Materials and methods

Host and *Wolbachia* genomes

The sources for the genomes and their hosts are given in Supplementary Table 1. For the *Wolbachia* that overlap with those in Vancaester and Blaxter (2023), we used the same genome ($N = 22$), an alternative complete genome ($N = 2$), or a higher quality genome ($N = 3$) (Supplementary Table 1). The latter were the circularized *wTei* and *wYak* genomes produced by Baiao et al. (2021), and a *wDal* genome from Scholz et al. (2020) that provided a higher quality assembly. We also include new genomes ($N = 10$) (Supplementary Table 1). Supplementary Table 1 also presents the sources for the draft nuclear, mitochondrial and *Wolbachia* genomes for *Nomada* bees, (*N. ferruginata*, (*N. panzeri*, (*N. flava*, (*N. leucophthalma*))), and 2 trios of filarial nematodes—(*O. volvulus*,

O. ochengi), *D. immitis*) and ((*B. malayi*, *B. pahangi*), *W. bancrofti*)—used for calibration. Data from *Nasonia* wasps, *N. longicornis* and *N. giraulti* include 9 single-copy nuclear regions (Supplementary Table 2 in Raychoudhury et al. 2009); and 2 mitochondrial genomic fragments: a 1,048 bp fragment of CO1 and a ~1,200 bp fragment that includes complete ATP6 and ATP8 gene sequences (presented in Oliveira et al. 2008). For the *Nasonia* host nuclear genomes, we estimated divergence using loci specified below; for their mitochondria, we used data from all 13 protein-coding loci.

Estimation of phylogenies and chronograms

Our estimates of all phylogenies and chronograms were obtained using RevBayes 1.1.1 (Hohna et al. 2016) and used only protein-coding loci. We describe our general methods here, whereas we report the data used for specific analyses in the following sections. For host nuclear genes, we partitioned by locus and codon position. For *Wolbachia* and host mtDNA, we partitioned only by codon position, with the exception of *Wolbachia* single-gene trees, which were unpartitioned. Phylograms were estimated using the substitution model GTR + I + $\Gamma(7,7)$, and chronograms were estimated using GTR + $\Gamma(7,7)$. We did not assess the relative adequacy of simpler substitution models because the results of Abadi et al. (2019) suggest that results from GTR + Γ are likely to be robust. (Analyses with a strict clock produced smaller divergence-time estimates and more relaxed clocks, e.g. $\Gamma(2,2)$, lose node resolution.)

We estimated relaxed-clock relative chronograms with the root age fixed to 1 and used the same birth-death prior as Turelli et al. (2018). Each partition had an independent rate multiplier with prior $\Gamma(1,1)$, as well as stationary frequencies and exchangeability rates drawn from flat, symmetrical Dirichlet distributions. The branch-rate prior was $\Gamma(7,7)$, normalized to a mean of 1 across all branches.

We ran phylogram analyses for 100,000 RevBayes generations and chronograms for 200,000 generations. Unlike MrBayes (Ronquist et al. 2012), in which “one generation” corresponds to a single proposed move in parameter space, 1 generation in RevBayes involves many proposals to accommodate model complexity (Hohna et al. 2017). For each analysis, we completed 4 independent RevBayes runs. Convergence for each run was assessed using Tracer (Rambaut et al. 2018). In all our analyses, the 4 runs converged to the same result. Nodes with posterior probability less than 0.95 were collapsed into polytomies.

Phylogenies and chronograms for hosts of *wMel*-like *Wolbachia*

Our estimates concerning host divergence rely on recent calibrations from the literature, as discussed in our Results. Our host species (presented in Table 1 & Supplementary Table 1) are extraordinarily diverse (see Results). For the drosophilids in Table 1, node ages and approximate support intervals presented in Fig. 1b and Table 2 were estimated from the fossil-calibrated chronogram of Suvorov et al. (2022), using Supplementary information as needed. For drosophilid species in Supplementary Table 1 that are not explicitly provided in Fig. 1 of Suvorov et al. (2022), we used the NCBI Taxonomy Browser to determine the closest relative(s) included in Suvorov et al. (2022) (Supplementary Table 1). The node ages and approximate support intervals not explicitly shown in their Fig. 1 were kindly provided by A. Suvorov (pers. comm.).

We used the Suvorov et al. (2022) calibrations for host divergence when possible. For species not included in Suvorov et al. (2022), we estimated divergence with a relative chronogram

including the species pair in question and a node present in Suvorov et al. (2022). Specifically, to estimate the *D. borealis*–*D. incompta* divergence time, we estimated a relative chronogram for *D. hydei*, *D. virilis*, *D. borealis*, and *D. incompta*, for which *D. hydei* is the outgroup, and an age estimate for the root is available from Suvorov et al. (2022). Similarly, to estimate the *D. malagassy*–*D. seguyi* and *D. sp. aff. chauvacae*–*D. bocqueti* divergences, we estimated a relative chronogram for (*D. jambulina*, ((*D. bocqueti*, *D. sp. aff. chauvacae*), (*D. malagassy*, *D. seguyi*))) with an age estimate for the root from Suvorov et al. (2022). For these chronograms, we estimated relative divergence using relaxed-clock relative chronograms under the GTR + $\Gamma(7,7)$ model of molecular evolution, following the methods in Conner et al. (2021), except that we partitioned by both gene and codon position. For these trees, we used the coding sequences for the 20 nuclear genes used in the analyses of Turelli et al. (2018): *aconitase*, *aldolase*, *bicoid*, *ebony*, *Enolase*, *esc*, *g6pdh*, *GlyP*, *GlyS*, *ninaE*, *pepck*, *Pgi*, *Pgm1*, *pic*, *ptc*, *Tpi*, *Transaldolase*, *white*, *wingless*, and *yellow*. Coding sequences for these genes in *D. melanogaster* were obtained from FlyBase. We used tBLASTn to identify homologs in the other genome assemblies. The sequences were aligned with MAFFT v. 7 (Katoh and Standley 2013) and trimmed of introns using the *D. melanogaster* sequences as a guide. The resulting sequences were visually inspected for gaps, which only occurred in multiples of 3 and did not create frameshifts.

Estimation of mtDNA divergence times requires a plausible range of mtDNA substitution rates. Because these rates may decline with species divergence (Ho et al. 2005), we computed mtDNA third-position substitution rates for 5 *Drosophila* species pairs that differ in their estimated divergence times (Drosophila 12 Genomes Consortium et al. 2007; Suvorov et al. 2022): *D. melanogaster*–*D. simulans* (3.62 MY, 95% CI: 2.92 to 4.4 MY), *D. erecta*–*D. yakuba* (5.47 MY, 95% CI: 4.59 to 6.38 MY), *D. mojavensis*–*D. virilis* (24.34 MY, 95% CI: 21.94 to 26.88 MY), *D. pseudoobscura*–*D. suzukii* (32.78 MY, 95% CI: 29.77 to 35.76 MY), and *D. grimshawi*–*D. willistoni* (46.84 MYA, 95% CI: 43.85 to 49.85 MY). For each of these 10 species, coding sequences for the 13 protein-coding mitochondrial genes in the inbred reference strains were obtained from FlyBase. They were aligned with MAFFT v.7 and concatenated. We counted nucleotide differences and applied the Jukes-Cantor correction for multiple substitutions. mtDNA third-position substitution rates were computed using the Jukes-Cantor estimates for the number of substitutions and fossil-calibrated divergence time estimates of Suvorov et al. (2022). Our possible examples of introgressive *Wolbachia* transfer involve relatively recently diverged pairs (Results). Thus, we averaged the point estimates for third-position mtDNA substitution rates for the 2 most recently diverged *Drosophila* pairs, *D. melanogaster*–*D. simulans* and *D. erecta*–*D. yakuba*. We used the fastest and slowest substitution-rate estimates across both sets of credible intervals to produce a conservative plausible range of mtDNA third-position substitution rates. We used these rate estimates and the observed mtDNA third-position differences between species pairs to estimate their mtDNA divergence times and plausible ranges.

Chronograms for hosts used to calibrate *Wolbachia* divergence rates

Nasonia. Raychoudhury et al. (2009) used a *Drosophila*-based calibration (Tamura et al. 2004) for synonymous divergence of host nuclear genes to estimate that *N. giraulti* and *N. longicornis* diverged about 0.51 MYA.

Filarial nematodes. We consider 2 trios of filarial nematodes, each with a pair of congeners and an outgroup containing a

Wolbachia variant belonging to the same supergroup as those in the congeners: ((*O. volvulus*, *O. ochengi*), *D. immitis*) and ((*B. malayi*, *B. pahangi*), *W. bancrofti*). For both trios, Qing et al. (2025) estimated the time to the MRCA of the 3 species. For each trio, we used their crown-age estimates and then interpolated internal node ages from multilocus-based relative chronograms. For each of these 6 nematode species, we obtained 33,099 bp of nuclear data (Supplementary Table 1). Our host relative chronograms were estimated using data from the 20 single-copy nuclear loci: *aldolase*, *ddb*, *fdh*, *g6pdh*, *galactokinase*, *glyp*, *glys*, *idh*, *ldh*, *lias*, *mes*, *pgi*, *pgm*, *pisy*, *ptc*, *rpi*, *sqv*, *tag*, *transaldolase*, and *vem*. These metabolic housekeeping genes exist as single-copy homologs in each of the 6 species. Homologs were aligned with MAFFT v. 7 (Katoh and Standley 2013). We visually inspected the alignments for gaps, which occurred only in multiples of 3 and produced no frameshifts in the alignments. In assessing our *Wolbachia* calibrations, we also analyzed host mtDNA to cross-calibrate divergence with the nuclear and *Wolbachia* loci.

Nomada. We reanalyzed Gerth and Bleidorn's (2016) proposed example of codivergence of *Wolbachia* within the 4-species clade of *Nomada* bees, (*N. ferruginata*, (*N. panzeri*, (*N. flava*, *N. leucophthalma*))). Our analyses depend on comparing rates of divergence for the *Nomada* nuclear genomes, *Nomada* mtDNA, and their associated *Wolbachia* (Supplementary Table 1). For the crown age of the 4-species clade, we used the estimate and 95% credible interval, 2.14 (1.16 to 3.37 MY), obtained from Fig. 3 and p. 6 of Gerth and Bleidorn (2016). Internal node ages were estimated from relative chronograms derived from 36,402 bp of nuclear protein-coding sequence from 20 loci from each of the 4 species: *aconitase*, *arl2*, *beta-hexosaminidase*, *citrate synthase*, *cytidine deaminase*, *esc*, *g6pdh*, *galactokinase*, *glys*, *houki*, *idh*, *las*, *pic*, *pis*, *protein kinase JNK*, *rpi*, *sgl*, *shmt*, *transaldolase*, and *UBFD1*. The point estimates for the ages of the internal nodes were based on the crown age estimate of 2.14 MY. Support intervals for internal node ages are based on the 95% credible intervals estimated from the relative chronogram.

Wolbachia assembly and phylogeny estimation

For the 10 genomes novel to this study, we trimmed the reads using Sickle v. 1.3 (Joshi and Fass 2011) and assembled them with ABySS v. 2.2.3 (Jackman et al. 2017) with Kmer values of 51, 61, ..., 91. From these host assemblies, scaffolds with best nucleotide BLAST matches to known *Wolbachia* sequences, with E-values less than 10^{-10} , were extracted as the *Wolbachia* assembly. For each host, the best *Wolbachia* assembly (fewest scaffolds and highest N50) was retained. To assess the quality of these draft assemblies, we used BUSCO v. 3.0.0 (Simao et al. 2015) to search for homologs of the near-universal, single-copy genes in the BUSCO proteobacteria database (<https://busco.ezlab.org/>). As controls, we performed the same search using the reference genomes for *wRi*, *wAu*, *wMel*, *wHa*, and *wNo*.

To identify genes for the phylogenetic analyses, all our *wMel*-like and *wRi*-like *Wolbachia* genomes (Supplementary Table 1) were annotated with Prokka 1.14.5 (Seemann 2014). In cases where genes in different genomes match the same reference gene according to Prokka, we considered them homologous. To avoid pseudogenes, paralogs and frameshifts, we used only genes present in a single copy in each genome and required that at most 1 draft genome had gaps in the alignment; the deletions all were in frame, involving sets of 3 base pairs. Genes were identified as single-copy if they uniquely matched a bacterial reference gene identified by Prokka. They were then aligned with MAFFT v. 7 (Katoh and Standley 2013). We estimated 3 separate phylogenies.

For the set of the 20 *wMel*-like genomes, 438 genes, a total of 409,848 bp, met these criteria. For the 20 *wMel*-like genomes plus the 8 *wRi*-likes, 346 genes, a total of 310,977 bp, met these criteria. For the 8 *wRi*-like genomes above and here, we used the same data set used in [Turelli et al. \(2018\)](#), which includes 525 genes, a total of 506,307 bp. These “core” genes within the *wMel*-like and *wRi*-like clades show little evidence for intragenic recombination or horizontal acquisition (see Results).

Assessing recombination and the mosaic structure of closely related *Wolbachia* genomes

In the following subsection, we describe single-locus tests for recombination that search for discordance between phylogenies associated with different segments of individual genes. In addition to looking for recombination within coding loci, we consider a range of other tests to assess the accuracy of our phylograms and chronograms. We determine whether individual *Wolbachia* loci support phylogenies different from those estimated from concatenating data across loci and assess possible causes of discordance. We also determine the loci that show the highest levels of allelic variation across various sets of reference strains. These most-variable loci are plausible candidates for horizontal acquisition of alleles from outside the reference set. We examined patterns of pairwise differences among alleles, looking for outliers suggesting horizontal acquisition.

Detecting recombination within genes among genomes. To test for intralocus recombination across *Wolbachia* genomes, we used a genetic algorithm for recombination detection (GARD) ([Kosakovsky Pond et al. 2006](#)), plus all 3 statistical methods implemented in PhiPack: the Pairwise Homoplasy Index (PHI), Maximum Chi Square (χ^2), and Neighbor Similarity Score (NSS) tests ([Bruen et al. 2006](#)). The PHI test detects recombination by evaluating phylogenetic incompatibility between sites, χ^2 divides the alignment into windows and assesses differences in the distribution of substitutions to identify potential breakpoints, and NSS measures similarity along the sequence to identify reductions indicative of a breakpoint. GARD is widely implemented, including in analyses of intralocus *Wolbachia* recombination ([Wang et al. 2020](#)). We focused our analyses on single-copy genes at least 300 bp in length, with minimum recombination segments of 100 bp. The 300 bp criterion excludes a small number of the single-copy genes within each set described above. We set GARD to detect a maximum of 2 potential breakpoints, and used the GTR + Γ model with 4 rate categories. For our 20 *wMel*-like variants, 411 genes (403,206 bp) met our length criteria.

We used these same analyses to test for recombination in 3 additional sets of *Wolbachia* with different levels of sequence divergence. We analyzed the 8 *wRi*-like *Wolbachia* presented in [Fig. 1](#) in [Turelli et al. \(2018\)](#). 525 genes (506,307 bp) met our criteria. We also examined the 7 more diverged supergroup A *Wolbachia* genomes introduced above: *wMel*, *wAu*, *wRi*, *wTri*, *wHa*, *wBic* and *wBar*. For this set, 401 genes (398,602 bp) met our criteria. Finally, we considered more anciently diverged *Wolbachia* in filarial nematodes relevant to our calibrations. We analyzed 4 supergroup D *Wolbachia* with the following phylogeny: (*wWb* in *W. bancrofti*, (*wBm* in *B. malayi*, *wBp* in *B. pahangi*)) and the outgroup *Wolbachia* (*wLs*) associated with *Litomosoides sigmodontis*. For these 4, 356 genes (349,602 bp) met our criteria. For our 3 supergroup C *Wolbachia* trio (*wOv* in *O. volvulus*, *wOo* in *O. ochengi*, and *wDimm* in *D. immitis*), we could not find an additional supergroup C *Wolbachia* outgroup. This precludes analysis of within-gene discordance for supergroup C because there is only one unrooted phylogeny for 3 taxa.

We assessed the reliability of the GARD recombination results for the 7 supergroup A *Wolbachia* by visually inspecting conflicting trees inferred by GARD and the data underlying them. As discussed in our Results, GARD makes artifactual recombination calls when there is insufficient sequence variation to produce reliable phylograms ([Supplementary Information](#)).

We next describe several independent analyses that provide strong support for the bifurcating evolution of major portions of the *Wolbachia* genome, including for the 7 supergroup A *Wolbachia*.

Discordance of gene trees. As an alternative assay for mosaic genome structure, we looked for variation among the unrooted topologies inferred from 442 single-copy protein-coding loci (total 413,115 bp) from the focal 7 supergroup A *Wolbachia*. For each gene, we attempted to infer an unrooted tree for all 7 *Wolbachia*. For illustration, we highlighted the subset of 52 loci that produced fully resolved trees, with posterior probabilities > 0.95 for each node, and compared them to the consensus topology estimated from the concatenated sequence data (52 genes; 72,027 bp). To assess discordance, we estimated an unrooted species tree from all 442 gene trees using weighted ASTRAL v1.23 ([Zhang and Mirarab 2022](#)), and used IQ-TREE v. 3.0.1 to calculate gene concordance factors (gCF) on the reference topology ([Wong et al. 2025](#)). Using 50%, 60%, 80%, and 95% bootstrap thresholds, we collapsed nodes with low support in the gene trees and compared results across thresholds. Default settings were used for ASTRAL and IQ-TREE analyses.

We applied the same discordance analysis to 7 nematode-derived *Wolbachia* genomes, the 4 supergroup D genomes used for our intralocus recombination assay above and 3 supergroup C genomes used for our calibration. This tests for horizontal transmission of fragments across these genomes.

Polymorphism and pairwise allele-difference assays for horizontal transmission. A limitation of gene-tree discordance analyses is that they require sufficient differences among the alleles to produce a fully resolved phylogeny. If only 1 genome horizontally acquired a distantly related allele (or allele fragment) at a specific locus, the remaining alleles may not be sufficiently different to provide full phylogenetic resolution. As an alternative assay for horizontal transmission, we ranked the single-copy loci by their levels of polymorphism, measured as the fraction of variable nucleotide sites, across our *wMel*-like and *wRi*-like *Wolbachia* and across our 7 more diverse supergroup A *Wolbachia*. The most polymorphic loci are candidates for having horizontally acquired alleles. To account for variation in rates of molecular evolution across loci, we directly examined pairwise differences between all alleles, looking for outliers as candidates for horizontal acquisition.

Based on a preliminary examination of pairwise differences between alleles at single-copy loci, we applied a heuristic criterion for identifying plausible examples of horizontally acquired alleles. For individual loci within a specific set of *Wolbachia*, such as our *wMel*-like and *wRi*-like variants, we categorized an allele as a 2-fold outlier if it differs from all other alleles in the set at twice as many nucleotides as any of the remaining alleles differ from one another. This heuristic criterion is obviously most easily satisfied for closely related *Wolbachia*. For loci with 2-fold outliers, we examined the flanking loci and mapped the nucleotide differences within the locus to assess the size of the fragment that was plausibly horizontally acquired.

Calibrating rates of *Wolbachia* divergence

We produce new calibrations based on codivergence of *Wolbachia* and filarial nematode hosts. [Raychoudhury et al. \(2009\)](#) pioneered

codivergence calibrations. Before describing our new filarial nematode analyses and our reanalysis of data from *Nomada* bees, we describe our reanalysis of the *Nasonia* calibration from Raychoudhury et al. (2009).

Nasonia calibration. To determine whether any of the 11 *Wolbachia* variants found in 4 *Nasonia* species were plausible candidates for cladogenic acquisition, Raychoudhury et al. (2009) compared molecular-clock-based estimates for host divergence times (described above) with molecular-clock-based divergence-time estimates for particular pairs of their *Wolbachia*. To produce a plausible divergence time for the *Wolbachia* wNgirB and wNlonB1, they used a synonymous bacterial divergence rate (double the substitution rate) of 9×10^{-9} . This produces an age estimate of 0.41 MYA for the MRCA of wNgirB and wNlonB1, comparable to that for their hosts. This rough agreement based on 2 plausible molecular clocks is the key evidence supporting *Wolbachia*-host codivergence in these *Nasonia*. To understand the implications of this hypothesized *Nasonia*-*Wolbachia* codivergence for *Wolbachia* substitution rates, we averaged the eukaryotic-based and prokaryotic-based divergence-time estimates to approximate the age of the MRCA of both pairs (*N. giraulti*, *N. longicornis*) and (wNgirB, wNlonB1) as about 0.46 MYA. We reanalyzed the raw data available from Raychoudhury et al. (2009) for these host nuclear genomes and these two *Wolbachia* variants to produce the divergence-time estimates. We also compared the hosts' mtDNA to compare with other examples of putative cladogenic *Wolbachia* acquisition.

Filarial nematode calibrations. For each species of the 2 host trios—(*O. volvulus*, *O. ochengi*), *D. immitis*) and (*B. malayi*, *B. pahangi*), *W. bancrofti*)—we extracted 598,257 bp from single-copy *Wolbachia* loci. To look at relative rates of *Wolbachia* and host mtDNA divergences, we also extracted the 13 protein-coding mtDNA genes, with a total of 10,361–10,373 bp. To estimate pairwise divergences, we aligned the genes, counted nucleotide differences and applied the Jukes-Cantor correction for multiple substitutions. When computing rates associated with each outgroup, we averaged the 2 sequence divergence estimates from each sister to the outgroup. We estimated rates and their support intervals using our point estimates and support intervals for host divergence times.

Nomada calibrations. The new nematode data allow us to critically examine the putative codivergence of *Wolbachia* with hosts in the 4-species clade of *Nomada* bees (*N. ferruginata*, (*N. panzeri*, (*N. flava*, *N. leucophthalma*))) (Gerth and Bleidorn 2016). To assess the implications and plausibility of *Wolbachia*-host codivergence throughout this clade, we compared rates of *Wolbachia* sequence divergence to the divergence of host mtDNA (using data kindly provided by M. Gerth and C. Bleidorn) and nuclear loci. Our analyses used the 13 protein-coding gene sequences, a total of 9,249 bp. We analyzed 613,605 bp of *Wolbachia* protein-coding sequence from single-copy loci. *Wolbachia* substitution rates were estimated as for the nematodes and *Nasonia*. We estimated separate *Wolbachia* rates from each node of the *Nomada* tree.

Wolbachia chronograms. We estimated chronograms for 3 datasets—wMel-like clade, wRi-like clade (Turelli et al. 2018), and both clades combined. We first estimated a relaxed-clock relative chronogram with the root age fixed to 1. There were not enough substitutions to partition by gene as well as codon position. In Turelli et al. (2018), we converted the relative chronogram into an absolute chronogram using the rate prior $\Gamma(7,7) \times 6.87 \times 10^{-9}$ substitutions per third-position site per year, derived from the mutation-based posterior distribution estimated by Richardson et al. (2012). The mean assumes 10 generations per year. Absolute branch lengths (times) were calculated as the relative

branch length times the third-position rate multiplier estimated from the substitutions per third-position site per year. Here, we consider alternative calibrations for the third-position rate, based on putative examples of cladogenic *Wolbachia* transmission in *Nasonia* and *Nomada*.

Turelli et al. (2018) used the Γ distribution to approximate the variation in the mutation-rate estimate obtained by Richardson et al. (2012). Although $\Gamma(7,7)$ approximated the per-generation variance estimate, Turelli et al. (2018) used it to approximate the per-year variance. Because Turelli et al. (2018) also assumed 10 generations per year, the variance per year was underestimated in this analysis. Here we use a different approach to approximating the uncertainty in substitutions per third-position site per year and explore the consequences of incorporating that uncertainty in our divergence-time estimates. We estimate plausible ranges of substitution rates for different subsets of *Nomada* host-*Wolbachia* comparisons based on estimated ranges for host divergence times (Table 4 in Results). The alternative calibrations of *Wolbachia* divergence rates produce a wide range of plausible divergence times, as illustrated in Tables 5 and 7. The variation among estimates based on alternative calibrations probably better represents the uncertainty of divergence-time estimates than do the probabilistic support intervals produced associated by individual calibrations.

For their *Nasonia* analyses, Raychoudhury et al. (2009) used widely applied eukaryotic and bacterial molecular clock calibrations to separately estimate the divergence times for hosts *N. longicornis* and *N. giraulti* and the *Wolbachia* inferred to be cladogenically acquired, wNlonB1 and wNgirB. We averaged those estimated divergence times for hosts and their *Wolbachia* to produce a divergence-time estimate of 0.46 MY, which produces an average third-site *Wolbachia* substitution rate of 2.2×10^{-9} substitutions per third site per year over the 4,486 bp of their *Wolbachia* DNA sequence data (Table 4 above).

For their *Nomada* pairs with plausible cladogenic *Wolbachia* transmission, Gerth and Bleidorn (2016) used fossil-based estimates of the host divergence times. Specifically, they estimated the crown age of the clade (*N. ferruginata*, (*N. panzeri*, (*N. flava*, *N. leucophthalma*))) as 2.14 MY, with 95% credible interval (1.16, 3.37) MY. As described above, we used additional single-copy nuclear data from these 4 species to more accurately estimate the relative ages of the internal nodes of this clade. We then converted the relative chronogram to an absolute chronogram by assuming that the crown age is 2.14 MY. We used the *Wolbachia* divergence rates between the hosts to estimate mean substitution rates.

Support intervals for the rates were approximated based on the support intervals for the host divergence times. The data-motivated models used are described in our Results. To illustrate the consequences of approximating uncertainty in the individual *Nomada* rate estimates, we compare divergence-time estimates and their support intervals produced by analyses assuming constant vs stochastically distributed divergence-rate priors. That uncertainty, associated with individual calibrations, is contrasted with the larger effects associated with alternative rate calibrations (Table 5).

CI phenotyping

While estimates of CI intensity exist for many of the *Wolbachia*-host pairs in our study, we tested for CI in 2 previously uncharacterized *D. montium* subgroup host species that carry wMel-like variants: wSeg in *D. seguyi* and wBocq in *D. bocqueti*. We secured *D. seguyi* and *D. bocqueti* stocks carrying wMel-like

Wolbachia from the National *Drosophila* Species Stock Center. To generate lines of both species without *Wolbachia*, we exposed *Wolbachia*-carrying lines to tetracycline-supplemented (0.03%) cornmeal medium (Cooper and Shropshire 2024a) for 3 generations. We initially confirmed the absence of *Wolbachia* in the treated flies using PCR within 2 generations of tetracycline treatment using primers for the *Wolbachia*-specific *wsp* gene (Braig et al. 1998; Baldo et al. 2005) and a second reaction for the arthropod-specific 28S rDNA as a host control (Nice et al. 2009; Cooper and Shropshire 2024b). Our PCR thermal profile began with 3 min at 94°C, followed by 34 rounds of 30 s at 94°C, 30 s at 55°C, and 1 min and 15 s at 72°C. The profile finished with a single round of 8 min at 72°C. We visualized PCR products using 1% agarose gels that included a molecular weight ladder. The stocks were maintained and experiments were conducted in an incubator at 25°C. Tetracycline-treated stocks were given at least 4 generations to recover prior to our CI experiments, and subsequent PCR-based tests confirmed that these stocks remained *Wolbachia*-free. All stocks were maintained under standard conditions using established protocols (Hartman et al. 2024; Wheeler et al. 2024).

We reciprocally crossed *Wolbachia*-carrying *D. seoyui* and *D. bocqueti* lines to their tetracycline-treated conspecifics and estimated the egg hatch frequencies from potentially incompatible crosses between females without *Wolbachia* and males with *Wolbachia* (denoted IC) and from the reciprocal compatible cross (CC) between females with and males without *Wolbachia*. With cytoplasmic incompatibility, we expect lower egg hatch from IC crosses than from CC crosses. Virgins were collected from each line and placed into holding vials for 48 h. We set up each IC and CC cross with 1 female and 1 male in a vial containing a small spoon with cornmeal medium and yeast paste for 24 h (Shropshire 2025). Males and females were 2 days old at the beginning of these experiments. Each pair was transferred to a fresh vial every 24 h for 5 days. We counted the number of eggs laid at the time the adults were transferred to new vials and the number of eggs that hatched were scored after an additional 24 h. The data analyzed were hatch proportions for crosses across the 5-day period. To control for cases where females may not have been inseminated, we excluded crosses that produced fewer than 10 eggs. We used one-sided Wilcoxon tests to determine whether IC crosses produce lower egg hatch proportions than do CC crosses.

Extracting sr alleles to type Wovirus

We used sr alleles to type *Woviruses* in *wMel*-like and *wRi*-like genomes, focusing on *Woviruses* (i.e. sr1WO-sr3WO) known to contain *cifs* (Bordenstein and Bordenstein 2022). sr genes assist with phage genome integration into and excision from the bacterial chromosome. Their phylogeny corresponds to module synteny and genome integration, making sr alleles a useful *Wovirus* typing tool (sensu Bordenstein and Bordenstein 2022). Across all *wMel*-like and *wRi*-like *Wolbachia*, we used WOCauB3, WOvitA1, WOMelB, and WOFol2 sr alleles as queries in BLAST searches to characterize sr1WO, sr2WO, sr3WO, and sr4WO *Wovirus*, respectively. If the *Wolbachia* assembly clearly assigned an sr sequence to a phage, we typed the phage (sr1WO, sr2WO, sr3WO, or sr4WO). For all of these cases, sr sequences in the *Wolbachia* assemblies mapped to only 1 of the 4 query sequences.

Extracting cif sequences

For our set of *wMel*-like and *wRi*-like *Wolbachia*, we used BLAST to identify contigs with *cif* sequences with E-values less than 10^{-10} . Our search used the following reference genes to query and categorize *cif* sequences: Type 1 from *wMel* (*cif_{wMel}[T1]*), Type 2 from *wRi* in

D. simulans (*cif_{wRi}[T2]*), Type 3 from *wNo* in *D. simulans* (*cif_{wNo}[T3]*), Type 4 from *wPip* in *C. pipiens* (*cif_{wPip}[T4]*), Type 5 from *wStri* in *Laodelphax striatellus* (*cif_{wStri}[T5]*), Type 5 from *wTri* in *D. triauraria* (*cif_{wTri}[T5]*), and Type 5 from *wBor* in *D. borealis* (*cif_{wBor}[T5]*). These *cifs* are also classified based on predicted enzymatic domains of the *cifB* locus involved in CI induction (Oladipupo and Hochstrasser 2025). Type 1 *cifs* are referred to as *cid* loci based on a C-terminal deubiquitylase domain, Types 2 to 4 are referred to as *cin* loci based on predicted nuclease domains, and Type 5 *cifs* are referred to as *cnd* loci based on predicted nuclease and deubiquitylase domains (Beckmann et al. 2017; Chen et al. 2020). We use “Type” classification here, while acknowledging alternative proposals for naming these loci (Beckmann et al. 2019; Shropshire et al. 2019).

We assigned the *cif* sequences to Types based on similarity to the reference genes noted above. We describe specific *cifs* using up to 3 pieces of information in a subscript behind the gene name: strain, Type designation, and copy number. For example, *wRi* encodes 2 T1 *cif*A variants that we denote *cif_A_{wRi}[T1-1]* and *cif_A_{wRi}[T1-2]*. We used Geneious Prime to extract ORFs with blast homology to *cif* sequences for downstream analyses (Kearse et al. 2012). Among all *cif* sequences, only *cif_B_{wMal}[T1]* did not have a clear ORF; however, we did observe sequence homology in the region. To establish sequence integrity, we classified *cif* genes as intact or disrupted based on whether mutations either split the gene into multiple ORFs or truncate the protein, causing the loss of a functional domain.

Because recombination can bias phylogenetic inference (Bonneau et al. 2018; Cooper et al. 2019; Martinez et al. 2021), we tested for intragenic recombination using GARD (Kosakovsky Pond et al. 2006) and all 3 statistical methods implemented in PhiPack (Bruen et al. 2006). We tested for recombination segments with a minimum length of 100 bp, allowing no more than 2 breakpoints. GARD calls were visually assessed for reliability. When recombination was detected, we estimated maximum likelihood phylograms for each gene region using RAXML-NG v1.2.2 under the GTR+ Γ model with default settings, and performed 1,000 bootstrap replicates to assess node support.

Approximating cif-Wovirus associations

To approximate *cif* associations with *Woviruses*, we searched for contigs containing one or more sr and *cifA* alleles. We then quantified the distance between each *cifA/B* allele and the nearest sr. We evaluated the orientation of sr alleles relative to putatively associated *cifs* and characterized the number and type of genes between them based on NCBI annotation. In the Supplementary Information, we report several lines of evidence that support the plausible association of several *cif_A[T1]* alleles and sr alleles of sr3WO (sr3). Because they are the most common *cifs* and srs in our data, we estimated *cif_A[T1]* and sr3 phylograms for copies that are plausibly linked. The sequences were aligned with MAFFT v. 7 (Kato and Standley 2013). We again used GARD (Kosakovsky Pond et al. 2006) and all 3 statistical methods implemented in PhiPack (Bruen et al. 2006) to test for recombination in this set of *cif_A[T1]* alleles and in the set of associated sr3 alleles. Phylograms for sr3, and separately for GARD-identified partitions, were estimated with RevBayes v. 1.1.1. We used a GTR+ Γ +I model with 4 rate categories and no partitioning. We used a rate multiplier with prior $\Gamma(1,1)$ [i.e. Exp(1)], as well as stationary frequencies and exchangeability rates drawn from flat, symmetrical Dirichlet distributions [i.e. Dirichlet(1,1,1, . . .)]. The model used a uniform prior over all possible topologies. Branch lengths were drawn from a flat, symmetrical Dirichlet distribution and thus summed to 1. Four independent runs were performed, and

all converged to the same topology. For additional details on the priors and their justifications, refer to the study by Turelli et al. (2018).

Characterizing Cif protein structures

We ran HHPred on a Linux kernel to identify putative functional domains using the Pfam-A_v35 and SCOPe70_2.07 databases (Zimmermann et al. 2018). We considered only annotations with > 80% probability. If alternative annotations produced probability > 0.8, we selected the annotation with the highest probability. We used AlphaFold2 (Jumper et al. 2021) to predict Cif structures. Entries in the “reduced database” provided with AlphaFold prior to 5/10/22 were used to generate multiple sequence alignments (MSA) within AlphaFold. We generated 5 structures for each protein and performed amber relaxation to prevent unrealistic folding patterns. We sorted the 5 relaxed models by mean pLDDT and used the top result in other analyses. We visualized protein structures using PyMol 2.5.2 (Schrodinger LLC 2015) and aligned Cif_[T1] proteins relative to Cif_{wMel[T1]} for imaging using Cealign.

To investigate Cif protein sequence and structural similarity, we calculated pairwise amino-acid identity ID and template-modeling scores (TM) (Zhang and Skolnick 2004). We used Muscle 5.0 to generate an ensemble of 100 MSAs from CifA and CifB protein sequences (Edgar 2022). From the ensemble, we extracted the alignment with the highest confidence score for each column, and calculated pairwise amino-acid identity from this alignment. TM scores indicate the degree of similarity between 2 superimposed structures, with TM = 0 indicating no similarity, TM = 1 corresponding to complete congruence, and TM > 0.5 indicating a homologous structure (Xu and Zhang 2010). We generated TM scores for Cif proteins pairwise in PyMol using the psico module (Holder 2023). We performed each analysis twice, switching the reference and target proteins. The directionality of the structure impacts the TM score because the score is normalized to the length of the target protein. We used a Mann-Whitney U test in R to compare the TM scores from CifA and CifB, excluding those that were either truncated or positioned at the edge of a contig. We assessed the relationship between TM score and percent identity using a Spearman correlation.

Characterizing the selective pressures on cifs

Our tests for selection on cifs rely on the estimated normalized ratio of non-synonymous (d_N) to synonymous (d_S) substitutions. We focus our analyses on cifA/B_[T1] since Type 1 copies are particularly common among wMel-like and wRi-like *Wolbachia*. The ratio $\omega = d_N/d_S$ can be unstable on very short timescales, as the estimates fluctuate due to stochastic sampling effects and ω approaches its long-term expectation only asymptotically (Yang and Nielsen 2000; Kryazhimskiy and Plotkin 2008). On very long timescales, high d_S leads to saturation and reduces power (Anisimova et al. 2001). For our analyses, we selected putatively intact cifA/B_[T1] and disrupted cifA_[T1] copies that met $0.1 \leq d_S \leq 2.0$, which is established as a conservative and suitable range for analyses of ω (Yang and Bielawski 2000; Yang 2006).

The majority of putatively intact wMel-like and wRi-like cifA/B_[T1] copies are closely related to one another and distantly related to a clade of 4 copies found in wDal, wZta, wSpa, and wAra. We excluded wAra from our analyses because cifA/B_[T1] in wSpa and wAra are nearly identical, with only one difference in cifB_[T1] and 4 in cifA_[T1]. We used the cifA/B_[T1] copies from wDal, wZta, wSpa, and wMel for all analyses of intact cif_[T1] described below (pairwise d_S range: cifA_[T1] = 0.23 to 0.35, cifB_[T1] = 0.38 to 0.48). Relative to the wMel copies, all cif_[T1] copies observed in wRi-like genomes and the other wMel-like genomes were too closely

related for inclusion (see Fig. 5c). Our analyses focus on the 3 putatively functional cifB_[T1] domains, Nuc1, Nuc2 and Dub, involved in CI (Beckmann et al. 2017, 2021; LePage et al. 2017; Chen et al. 2019; Kaur et al. 2022), which we annotated on the codon alignment. We also annotated cifB_[T1] non-domain regions. Given an insufficient number of disrupted cifA_[T1] copies in our wMel-like and wRi-like genomes, we searched for additional copies in the data of Martinez et al. (2021). Our analyses included wBic from *D. bicornuta*, wStv from *D. sturtevantii*, and wNfla from *N. flava*, along with the disrupted wMel-like wAch copy (pairwise d_S range: 0.34 to 0.86). Results related to disrupted cif_[T1] copies are presented in the Supplementary Information. We used these data and multiple methods to compare selection on cifA and cifB, and separately, to compare selection on the 3 cifB_[T1] protein domains (Nuc1, Nuc2 and Dub) and associated non-domain cifB_[T1] regions.

Codeml analyses. We applied several likelihood-based models implemented in PAML using codeml (Yang 1997, 2007). We aligned our sets of sequences using TranslatorX's PRANK mode (Abascal et al. 2010). We also used a direct PRANK alignment with flags -F to skip insertions and -codon that uses an empirical codon model (Loytynoja 2014). The latter was used only to assess the sensitivity of a subset of results to alignment, and such cases are noted explicitly as they appear. In all cases, we visually inspected the alignments for frameshifts, finding none. For disrupted cifA_[T1], we first aligned with MAFFT v7 using default parameters, and then manually deleted any frameshift-causing insertions, and manually filled any frameshift-causing deletions with “N”. We then realigned sequences using the TranslatorX PRANK mode unless noted otherwise. All codon-based analyses used the Goldman-Yang codon model (GY94) (Goldman and Yang 1994) with F3 × 4 codon frequencies (CodonFreq = 2) (Nielsen and Yang 1998; Yang et al. 2000). For each analysis, we initiated model fits from multiple starting values for the transition/transversion rate parameter (κ) (0.5, 1.0, 2.0) and ω (0.25, 0.5, 0.75, 1.0) to account for potential sensitivity of likelihood maximization to initial values. For each sequence pair, we used the estimates of (κ , ω) that produced the highest log-likelihood across the 12 combinations of initial values. We first assessed ω using runmode = -2 for pairwise comparisons. We focus on the median ω , rather than the mean, because of our small sample size ($N = 6$).

Using paired contrasts, we tested whether $\omega_{\text{cifA}} < \omega_{\text{cifB}}$ using a one-sided Wilcoxon signed-rank test. We report the Hodges-Lehmann (HL) estimator (Hodges and Lehmann 1963) of the median difference with an exact 95% confidence interval obtained by inverting the signed-rank test (Hollander et al. 2015). Confidence intervals are obtained by identifying the set of parameter values that would not be rejected by the Wilcoxon signed-rank test, that is, by inverting the test (Lehmann 1975). To test whether $\omega_{\text{cifA-intact}} < \omega_{\text{cifA-disrupted}}$ using unpaired cifA_[T1] copies, we applied the Mann-Whitney test with a one-sided alternative and the corresponding 2-sample HL estimator that provides an exact 95% CI via rank-sum inversion. We also estimated nonparametric Bayesian effect sizes using the Bayesian bootstrap. The Bayesian bootstrap is analogous to the classical bootstrap, but instead of resampling observations with replacement, it assigns random weights to each observation drawn from a Dirichlet(1, ..., 1) distribution. This can be implemented by normalizing independent Exp(1) samples, which yields the same distribution of weights. For each contrast, we applied these random weights to compute the linearly interpolated weighted median of the pairwise differences (the 0.5 quantile of the weighted empirical CDF with linear interpolation). We repeated this procedure $B = 20,000$ times to obtain a posterior sample of median effects. From this posterior, we report the posterior mean of the median effect, the 95% equal-tailed credible interval, and the posterior probability of the hypothesized directional effect,

$P(\Delta < 0)$. These Bayesian summaries are reported alongside the Wilcoxon and Mann–Whitney results.

We also analyzed codon models on phylogenetic alignments to estimate lineage-specific and site-specific variation in ω . Because recombination can bias phylogenetic inference and has been observed for *cifs* (Bonneau et al. 2018; Cooper et al. 2019; Martinez et al. 2021), we first tested for intragenic recombination using GARD (Kosakovsky Pond et al. 2006), and all 3 statistical methods implemented in PhiPack (Bruen et al. 2006), as described in the *cif* extraction section above. When recombination was detected, we estimated maximum likelihood phylograms for each gene region using RAXML-NG v1.2.2 under the GTR + Γ model with default settings, and performed 1,000 bootstrap replicates to assess node support. We partitioned our data according to their respective unrooted topologies in all subsequent selection analyses. This produced highly supported 2-part partitions for intact *cifA*_[T1] and disrupted *cifA*_[T1].

We evaluated 6 models in codeml that differ in their treatment of variation in ω across lineages and across sites. The one-ratio model (M0) assumes a single ω across all sites and lineages, and the branch model allows ω to vary among lineages. We assigned each ingroup branch to its own category, while constraining the 2 branches associated with the outgroup to share a single ω . This treats the tree as effectively unrooted under the branch model. To test for heterogeneity among sites, we fit models M1a, M2a, M7, and M8, described below. The “nearly neutral” model (M1a) assigns sites to classes with $\omega < 1$ or $\omega = 1$, while the positive-selection model (M2a) adds a class with $\omega > 1$. Model M7 assumes a beta distribution of ω across sites, restricted to values between 0 and 1, while the beta plus positive-selection model (M8) adds an extra site class in which ω estimates exceed 1. Each recombination partition was analyzed under its GARD and RAXML-supported topology, and likelihood-ratio tests (LRTs) were used for partition-specific model comparisons.

For models that include a positive-selection class (M2a, M8), we obtained BEB posterior probabilities and reported sites with posterior $P \geq 0.95$, prioritizing sites concordant across M2a and M8 (Yang et al. 2005). We visually evaluated alignments for these sites, looking for evidence of multi-nucleotide mutations (MNMs) that may produce false signatures of positive selection. MNMs that occur simultaneously—e.g. due to error-prone DNA polymerases—violate codon substitution models that assume mutations occur independently (Venkat et al. 2018). For downstream site-specific analyses, we masked sites supported by BEB with evidence of MNMs. For comparisons of domains to non-domain regions within *cifB*_[T1] to test whether $\omega_{\text{Domain}} < \omega_{\text{Non-domain}}$, we summarized means per-site posterior ω by region from model M8 or M7 and their 95% bootstrap confidence intervals, resampling sites within each region 10,000 times. We assessed significance with a site-level label permutation test, treating sites as exchangeable. We computed the observed Δ ($\omega_{\text{Domain}} - \omega_{\text{Non-domain}}$) and generated an empirical null by permuting region labels 10,000 times. P values were obtained by comparing the observed statistic to the permutation-based null distribution. We computed 95% confidence intervals by bootstrapping the difference of means directly—resampling sites with replacement within each region 10,000 times—and taking percentile bounds. We report effect sizes, permutation P values, and 95% confidence intervals from codon-level bootstrap resampling within partitions.

SWAKK Analyses. We also used the Sliding Window Analysis of Ka and Ks (SWAKK) webserver to evaluate constraint on putatively intact *CifA*_[T1] and on the 3 *CifB*_[T1] domains (Liang et al. 2006). SWAKK generates an alignment of 2 nucleotide sequences and maps it onto a provided tertiary structure. The average amino acid width spans ~3.8 to 4.3 Å (Jumper et al. 2021), and SWAKK

calculates ω (i.e. Ka/Ks) using a 10 Å spherical sliding window across the reference structure. This enables tests for selection on regions within 3-dimensional (3D) structures that may be functionally important. We used AlphaFold structures for tertiary mapping with *CifA*/*B*_{wMel}[T1] as our reference, calculating ω for pairwise comparisons between the wMel proteins and each of the 3 other putatively intact *CifA*/*B*_[T1] proteins described above. We filtered SWAKK output to include only sites where $0.1 \leq Ks \leq 2.0$ and pLDDT ≥ 70 , which is considered a confident prediction for a given amino acid residue in an AlphaFold protein structure model (Jumper et al. 2021; Guo et al. 2022). Undefined entries and sentinel rows with $Ks = 99$ were excluded.

For each pair, we computed the difference of the median $\log(\omega + 10^{-6})$ between *CifA*_[T1] and *CifB*_[T1] (Δ), where $\Delta = \text{median}\{\log(\omega + 10^{-6})\}_{\text{CifA}} - \text{median}\{\log(\omega + 10^{-6})\}_{\text{CifB}}$. To summarize the effect size, we averaged the 3 Δ values to obtain the mean of the per-pair median differences on the log scale (Δ_p). We also computed the geometric-mean ratio $[\exp(\Delta_p)]$, where $\exp(\Delta_p) < 1$ indicates lower *CifA*_[T1] ω . To obtain 95% BCa (bias-corrected and accelerated) intervals for Δ_p , we used a non-circular, non-overlapping moving-block bootstrap (Kunsch 1989; Liu and Singh 1992) with $B = 5,000$ resamples. This approach preserves the short-range autocorrelation introduced by the 10 Å window. Within each pair and protein, we estimated the integrated autocorrelation time (IAT) of the ordered $\log(\omega + 10^{-6})$ series (Sokal 1997). Provisional block lengths were set to $2 \times \text{IAT}$ for each series, rounded to an integer, and truncated to the range 6 to 30 residues. The block length established for each pair (L) was the maximum of the *CifA*_[T1] and *CifB*_[T1] provisional lengths. For each bootstrap replicate, we resampled contiguous blocks with replacement until their combined length equaled or exceeded the original series length, truncating any excess to match the original size. The 3 Δ values were then recomputed and averaged to obtain a bootstrap Δ_p . BCa intervals used jackknife acceleration computed by leave-one-pair-out Δ_p . To obtain a one-sided P value for ($\Delta_p < 0$), we used a block label permutation within each pair that preserves local dependence (Good 2005; Winkler et al. 2014, 2015). We partitioned the ordered $\log(\omega + 10^{-6})$ series for *CifA*_[T1] and *CifB*_[T1] within each comparison into non-overlapping blocks of length L (same as in the bootstrap), pooled all blocks, and randomly reassigned blocks back to *CifA*_[T1] or *CifB*_[T1] independently for each label without replacement until the original series lengths were matched. For each permuted dataset, we recomputed Δ for each comparison and then averaged across comparisons to obtain Δ_p . The one-sided P -value for ($\Delta_p < 0$) is $(r + 1)/(B + 1)$, where r is the number of permuted Δ_p values $\leq$ the observed Δ_p , and $B = 10,000$ permutations.

To evaluate constraint on *CifB*_[T1] domains, we compared Nuc1, Nuc2, and Dub *CifB*_[T1] domains individually to pooled non-domain regions. For each domain, we computed $\Delta = \text{median}\{\log(\omega + 10^{-6})\}_{\text{domain}} - \text{median}\{\log(\omega + 10^{-6})\}_{\text{non-domain}}$. We averaged the 3 Δ values to compute Δ_p , with $\exp(\Delta_p)$ interpreted as the domain/non-domain ω ratio. All domain analyses used the same moving-block bootstrap/BCa framework and the same permutation logic as above, except L was chosen separately for domain and non-domain regions within each pair. Permutations reassembled both series independently from the pooled blocks, ensuring that a block can appear in both labels within a permutation. We excluded the wMel-wDal pair from the Dub analysis due to a paucity of domain windows.

Figure generation

We produced and/or edited figures in R, FigTree, Inkscape 1.1, Adobe Illustrator, and Keynote.

Data availability

The authors affirm that all data necessary for confirming the conclusions of this article are represented fully within the article and its tables and figures. Sources for all genomic data are provided in [Supplementary Table 1](#). Additional data are available at Dryad (doi:10.5061/dryad.g1jwstr6j), and code is available on GitHub (brandoncooper/calibrating-wolbachia). [Supplementary Movies](#) can be found at FigShare (<https://doi.org/10.25386/genetics.30446210>).

Supplemental material available at [GENETICS](#) online.

Acknowledgments

We thank Daniel Matute for directing us to data for *Zaprionus* species, Miles Zhang for comments related to the nematode data and for access to some results before publication, Mathieu Gaultier for pointing out an error in [Turelli et al. \(2018, Fig. 1\)](#) that is corrected in our Discussion, and A. Suvorov, E. Loreto, R. Unckless, K. Dyer, M. Gerth, and C. Bleidorn for providing data used in our analyses. Tim Wheeler provided support in the laboratory. Discussions with Andrew Kern improved our analyses, and Nitin Ravikanthachari and John Statz provided comments that improved a prior version of our manuscript. We thank three reviewers whose comments significantly improved our presentation.

Funding

Research was supported by National Science Foundation (NSF) CAREER (2145195) and National Institutes of Health (NIH) MIRA (R35GM124701) awards to BSC. Additional research support was provided by the University of Montana Genomics Core (UMGC) and Montana INBRE Data Science Core, which are funded by the National Institute of General Medical Sciences (P20GM103474), the Office of the Vice President for Research and Creative Scholarship at the University of Montana, and the M. J. Murdock Charitable Trust (including award 202324717 to BSC). An NSF Postdoctoral Research Fellowship in Biology (DBI-2010210) and start-up funding from Lehigh University also supported JDS. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF, UMGc, or the NIH.

Conflicts of interest

The authors declare no competing interests.

Author contributions

Conceptualization: JDS, MT, BSC; Data curation: WRC, DV, BSC; Formal analysis: JDS, WRC, DV, MT, BSC; Funding acquisition: BSC; Investigation: JDS, WRC, DV, BSC; Methodology: JDS, WRC, DV, MT, BSC; Project administration: JDS, MT, BSC; Resources: BSC; Software: WRC, DV, BSC; Supervision: MT, BSC; Validation: JDS, WRC, DV, MT, BSC; Visualization: JDS, DV, BSC; Writing—Original draft: JDS, MT, BSC; Writing—Review & editing: JDS, WRC, DV, AAH, MT, BSC.

Literature cited

Abadi S, Azouri D, Pupko T, Mayrose I. 2019. Model selection may not be a mandatory step for phylogeny reconstruction. *Nat Commun.* 10:934. <https://doi.org/10.1038/s41467-019-08822-w>.
Abascal F, Zardoya R, Telford MJ. 2010. Translatorex: multiple alignment of nucleotide sequences guided by amino acid translations.

Nucleic Acids Res. 38:W7–W13. <https://doi.org/10.1093/nar/gkq291>.
Adams KL et al. 2021. *Wolbachia* cifB induces cytoplasmic incompatibility in the malaria mosquito vector. *Nat Microbiol.* 6:1575–1582. <https://doi.org/10.1038/s41564-021-00998-6>.
Ahmed MZ, Breinholt JW, Kawahara AY. 2016. Evidence for common horizontal transmission of *Wolbachia* among butterflies and moths. *BMC Evol Biol.* 16:118. <https://doi.org/10.1186/s12862-016-0660-x>.
Ahmed MZ et al. 2015. The intracellular bacterium *Wolbachia* uses parasitoid wasps as phoretic vectors for efficient horizontal transmission. *PLoS Pathog.* 11:e1004672. <https://doi.org/10.1371/journal.ppat.1004672>.
Amoros J et al. 2025b. Diversity and spread of cytoplasmic incompatibility genes among maternally inherited symbionts. *PLoS Genet.* 21:e1011856. <https://doi.org/10.1371/journal.pgen.1011856>.
Amoros J et al. 2025a Oct 4. Genomic islands in *Wolbachia* prophages drive amplification and diversification of cytoplasmic incompatibility genes in *Culex pipiens*. *bioRxiv* 680192. <https://doi.org/10.1101/2025.10.03.680192>, preprint; not peer reviewed.
Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol.* 18:1585–1592. <https://doi.org/10.1093/oxfordjournals.molbev.a003945>.
Baiao GC, Janice J, Galinou M, Klasson L. 2021. Comparative genomics reveals factors associated with phenotypic expression of *Wolbachia*. *Genome Biol Evol.* 13:evab111. <https://doi.org/10.1093/gbe/evab111>.
Bailly-Bechet M et al. 2017. How long does *Wolbachia* remain on board? *Mol Biol Evol.* 34:1183–1193. <https://doi.org/10.1093/molbev/msx073>.
Baldo L, Bordenstein S, Wernegreen JJ, Werren JH. 2006. Widespread recombination throughout *Wolbachia* genomes. *Mol Biol Evol.* 23: 437–449. <https://doi.org/10.1093/molbev/msj049>.
Baldo L, Lo N, Werren JH. 2005. Mosaic nature of the *Wolbachia* surface protein. *J Bacteriol.* 187:5406–5418. <https://doi.org/10.1128/JB.187.15.5406-5418.2005>.
Bandi C, Anderson TJC, Genchi C, Blaxter ML. 1998. Phylogeny of *Wolbachia* in filarial nematodes. *Proc R Soc Lond B Biol Sci.* 265: 2407–2413. <https://doi.org/10.1098/rspb.1998.0591>.
Baumann P et al. 1995. Mutualistic associations of aphids and prokaryotes: biology of the genus *Buchnera*. *Appl Environ Microbiol.* 61: 1–7. <https://doi.org/10.1128/aem.61.1.1-7.1995>.
Beckmann JF et al. 2019. The toxin-antidote model of cytoplasmic incompatibility: genetics and evolutionary implications. *Trends Genet.* 35:175–185. <https://doi.org/10.1016/j.tig.2018.12.004>.
Beckmann JF, Ronau JA, Hochstrasser M. 2017. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. *Nat Microbiol.* 2:17007. <https://doi.org/10.1038/nmicrobiol.2017.7>.
Beckmann JF, Van Vaerenbergh K, Akwa DE, Cooper BS. 2021. A single mutation weakens symbiont-induced reproductive manipulation through reductions in deubiquitylation efficiency. *Proc Natl Acad Sci U S A.* 118:e2113271118. <https://doi.org/10.1073/pnas.2113271118>.
Begun DJ et al. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* 5: e310. <https://doi.org/10.1371/journal.pbio.0050310>.
Berg DE, Howe MM. 1989. Mobile DNA. American Society for Microbiology.
Bonneau M et al. 2018. *Culex pipiens* crossing type diversity is governed by an amplified and polymorphic operon of *Wolbachia*. *Nat Commun.* 9:319. <https://doi.org/10.1038/s41467-017-02749-w>.
Bordenstein SR, Bordenstein SR. 2022. Widespread phages of endosymbionts: phage WO genomics and the proposed taxonomic

- classification of Symbioviridae. *PLoS Genet.* 18:e1010227. <https://doi.org/10.1371/journal.pgen.1010227>.
- Bordenstein SR, Marshall ML, Fry AJ, Kim U, Wernegreen JJ. 2006. The tripartite associations between bacteriophage, *Wolbachia*, and arthropods. *PLoS Pathog.* 2:e43. <https://doi.org/10.1371/journal.ppat.0020043>.
- Bordenstein SR, Wernegreen JJ. 2004. Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Mol Biol Evol.* 21:1981–1991. <https://doi.org/10.1093/molbev/msh211>.
- Boyle L, O'Neill SL, Robertson HM, Karr TL. 1993. Interspecific and intraspecific horizontal transfer of *Wolbachia* in *Drosophila*. *Science.* 260:1796–1799. <https://doi.org/10.1126/science.8511587>.
- Braig HR, Zhou WG, Dobson SL, O'Neill SL. 1998. Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. *J Bacteriol.* 180:2373–2378. <https://doi.org/10.1128/JB.180.9.2373-2378.1998>.
- Brinkmann H, van der Giezen M, Zhou Y, Poncelin de Raucourt G, Philippe H. 2005. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. *Syst Biol.* 54:743–757. <https://doi.org/10.1080/10635150500234609>.
- Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics.* 172:2665–2681. <https://doi.org/10.1534/genetics.105.048975>.
- Caspi-Fluger A et al. 2012. Horizontal transmission of the insect symbiont *Rickettsia* is plant-mediated. *Proc R Soc Lond B Biol Sci.* 279:1791–1796. <https://doi.org/10.1098/rspb.2011.2095>.
- Cerveau N et al. 2015. Genomic context drives transcription of insertion sequences in the bacterial endosymbiont *Wolbachia wVulC*. *Gene.* 564:81–86. <https://doi.org/10.1016/j.gene.2015.03.044>.
- Cerveau N, Leclercq S, Leroy E, Bouchon D, Cordaux R. 2011. Short- and long-term evolutionary dynamics of bacterial insertion sequences: insights from *Wolbachia* endosymbionts. *Genome Biol Evol.* 3:1175–1186. <https://doi.org/10.1093/gbe/evr096>.
- Chafee ME, Funk DJ, Harrison RG, Bordenstein SR. 2010. Lateral phage transfer in obligate intracellular bacteria (*Wolbachia*): verification from natural populations. *Mol Biol Evol.* 27:501–505. <https://doi.org/10.1093/molbev/msp275>.
- Chen H, Ronau JA, Beckmann JF, Hochstrasser M. 2019. A *Wolbachia* nuclease and its binding partner provide a distinct mechanism for cytoplasmic incompatibility. *Proc Natl Acad Sci U S A.* 116:22314–22321. <https://doi.org/10.1073/pnas.1914571116>.
- Chen H, Zhang M, Hochstrasser M. 2020. The biochemistry of cytoplasmic incompatibility caused by endosymbiotic bacteria. *Genes (Basel).* 11:852. <https://doi.org/10.3390/genes11080852>.
- Chrostek E, Pelz-Stelinski K, Hurst GDD, Hughes GL. 2017. Horizontal transmission of intracellular insect symbionts via plants. *Front Microbiol.* 8:2237. <https://doi.org/10.3389/fmicb.2017.02237>.
- Chrostek E, Teixeira L. 2015. Mutualism breakdown by amplification of *Wolbachia* genes. *PLoS Biol.* 13:e1002065. <https://doi.org/10.1371/journal.pbio.1002065>.
- Comandatore F et al. 2013. Phylogenomics and analysis of shared genes suggest a single transition to mutualism in *Wolbachia* of nematodes. *Genome Biol Evol.* 5:1668–1674. <https://doi.org/10.1093/gbe/evt125>.
- Comeault AA, Venkat A, Matute DR. 2016. Correlated evolution of male and female reproductive traits drive a cascading effect of reinforcement in *Drosophila yakuba*. *Proc R Soc Lond B Biol Sci.* 283:20160730. <https://doi.org/10.1098/rspb.2016.0730>.
- Conner WR et al. 2017. Genome comparisons indicate recent transfer of wRi-like *Wolbachia* between sister species *Drosophila suzukii* and *D. subpulchrella*. *Ecol Evol.* 7:9391–9404. <https://doi.org/10.1002/ece3.3449>.
- Conner WR et al. 2021. A phylogeny for the *Drosophila montium* species group: a model clade for comparative analyses. *Mol Phylogenet Evol.* 158:107061. <https://doi.org/10.1016/j.ympev.2020.107061>.
- Cooper BS, Ginsberg PS, Turelli M, Matute DR. 2017. *Wolbachia* in the *Drosophila yakuba* complex: pervasive frequency variation and weak cytoplasmic incompatibility, but no apparent effect on reproductive isolation. *Genetics.* 205:333–351. <https://doi.org/10.1534/genetics.116.196238>.
- Cooper BS, Sedghifar A, Nash WT, Comeault AA, Matute DR. 2018. A maladaptive combination of traits contributes to the maintenance of a *Drosophila* hybrid zone. *Curr Biol.* 28:2940–2947.e6. <https://doi.org/10.1016/j.cub.2018.07.005>.
- Cooper BS, Shropshire JD. 2024a. Eradicating *Wolbachia*: a tetracycline-based approach for *Drosophila melanogaster* and other fruit flies. *protocols.io.* <https://doi.org/10.17504/protocols.io.dm6gp9rwpvzp/v1>
- Cooper BS, Shropshire JD. 2024b. PCR-based cytotyping to detect *Wolbachia* in insect tissues. *protocols.io.* <https://doi.org/10.17504/protocols.io.3b4y49yx8go5/v1>
- Cooper BS, Vanderpool D, Conner WR, Matute DR, Turelli M. 2019. *Wolbachia* acquisition by *Drosophila yakuba*-clade hosts and transfer of incompatibility loci between distantly related *Wolbachia*. *Genetics.* 212:1399–1419. <https://doi.org/10.1534/genetics.119.302349>.
- Cordaux R, Michel-Salzat A, Bouchon D. 2001. *Wolbachia* infection in crustaceans: novel hosts and potential routes for horizontal transmission. *J Evol Biol.* 14:237–243. <https://doi.org/10.1046/j.1420-9101.2001.00279.x>.
- Cordaux R et al. 2008. Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. *Mol Biol Evol.* 25:1889–1896. <https://doi.org/10.1093/molbev/msn134>.
- Coyne JA, Orr HA. 1997. “Patterns of speciation in *Drosophila*” revisited. *Evolution.* 51:295–303. <https://doi.org/10.1111/j.1558-5646.1997.tb02412.x>.
- Coyne JA, Orr HA. 2004. Speciation. Sinauer Associates.
- Darby AC et al. 2012. Analysis of gene expression from the *Wolbachia* genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. *Genome Res.* 22:2467–2477. <https://doi.org/10.1101/gr.138420.112>.
- Deehan M, Lin W, Blum B, Emili A, Frydman H. 2021. Intracellular density of *Wolbachia* is mediated by host autophagy and the bacterial cytoplasmic incompatibility gene *cifB* in a cell type-dependent manner in *Drosophila melanogaster*. *mBio.* 12:e02205-20. <https://doi.org/10.1128/mBio.02205-20>.
- Drosophila* 12 Genomes Consortium et al. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature.* 450:203–218. <https://doi.org/10.1038/nature06341>.
- Duron O et al. 2005. Transposable element polymorphism of *Wolbachia* in the mosquito *Culex pipiens*: evidence of genetic diversity, superinfection and recombination. *Mol Ecol.* 14:1561–1573. <https://doi.org/10.1111/j.1365-294X.2005.02495.x>.
- Dyson EA, Kamath MK, Hurst GDD. 2002. *Wolbachia* infection associated with all-female broods in *Hypolimnas bolina* (Lepidoptera: Nymphalidae): evidence for horizontal transmission of a butterfly male killer. *Heredity (Edinb).* 88:166–171. <https://doi.org/10.1038/sj.hdy.6800021>.
- Edgar RC. 2022. Muscle5: high-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat Commun.* 13:6968. <https://doi.org/10.1038/s41467-022-34630-w>.
- Edwards SV. 2009. Is a new and general theory of molecular systematics emerging? *Evolution.* 63:1–19. <https://doi.org/10.1111/j.1558-5646.2008.00549.x>.
- Ellegaard KM, Klasson L, Naslund K, Bourtzis K, Andersson SGE. 2013. Comparative genomics of *Wolbachia* and the bacterial

- species concept. *PLoS Genet.* 9:e1003381. <https://doi.org/10.1371/journal.pgen.1003381>.
- Fenn K, Blaxter M. 2004. Are filarial nematode *Wolbachia* obligate mutualist symbionts? *Trends Ecol Evol.* 19:163–166. <https://doi.org/10.1016/j.tree.2004.01.002>.
- Ferri E et al. 2011. New insights into the evolution of *Wolbachia* infections in filarial nematodes inferred from a large range of screened species. *PLoS One.* 6:e20843. <https://doi.org/10.1371/journal.pone.0020843>.
- Gao S, Ren YS, Su CY, Zhu DH. 2022. High levels of multiple phage WO infections and its evolutionary dynamics associated with *Wolbachia*-infected butterflies. *Front Microbiol.* 13:865227. <https://doi.org/10.3389/fmicb.2022.865227>.
- Gavotte L et al. 2007. A survey of the bacteriophage WO in the endosymbiotic bacteria *Wolbachia*. *Mol Biol Evol.* 24:427–435. <https://doi.org/10.1093/molbev/msl171>.
- Gehrler L, Vorburger C. 2012. Parasitoids as vectors of facultative bacterial endosymbionts in aphids. *Biol Lett.* 8:613–615. <https://doi.org/10.1098/rsbl.2012.0144>.
- Gerth M, Bleidorn C. 2016. Comparative genomics provides a time-frame for *Wolbachia* evolution and exposes a recent biotin synthesis operon transfer. *Nat Microbiol.* 2:16241. <https://doi.org/10.1038/nmicrobiol.2016.241>.
- Gerth M, Roethe J, Bleidorn C. 2013. Tracing horizontal *Wolbachia* movements among bees (*Anthophila*): a combined approach using multilocus sequence typing data and host phylogeny. *Mol Ecol.* 22:6149–6162. <https://doi.org/10.1111/mec.12549>.
- Ghedini E et al. 2007. Draft genome of the filarial nematode parasite *Brugia malayi*. *Science.* 317:1756–1760. <https://doi.org/10.1126/science.1145406>.
- Gillespie JH, Langley CH. 1979. Are evolutionary rates really variable? *J Mol Evol.* 13:27–34. <https://doi.org/10.1007/BF01732751>.
- Gillespie JJ et al. 2018. A tangled web: origins of reproductive parasitism. *Genome Biol Evol.* 10:2292–2309. <https://doi.org/10.1093/gbe/evy159>.
- Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol.* 11:725–736. <https://doi.org/10.1093/oxfordjournals.molbev.a040153>.
- Gomes-de-Sa S et al. 2022. De novo assembly of the *Dirofilaria immitis* genome by long-read Nanopore-based sequencing technology on an adult worm from a canine cardiopulmonary dirofilariasis case. *Animals (Basel).* 12:1342. <https://doi.org/10.3390/ani12111342>.
- Gong JT, Li TP, Wang MK, Hong XY. 2023. *Wolbachia*-based strategies for control of agricultural pests. *Curr Opin Insect Sci.* 57:101039. <https://doi.org/10.1016/j.cois.2023.101039>.
- Good P. 2005. Permutation, parametric and bootstrap tests of hypotheses. Springer-Verlag. (Springer Series in Statistics). <https://doi.org/10.1007/b138696>.
- Gu X et al. 2022. A wMel *Wolbachia* variant in *Aedes aegypti* from field-collected *Drosophila melanogaster* with increased phenotypic stability under heat stress. *Environ Microbiol.* 24:2119–2135. <https://doi.org/10.1111/1462-2920.15966>.
- Guo HB et al. 2022. AlphaFold2 models indicate that protein sequence determines both structure and dynamics. *Sci Rep.* 12:10696. <https://doi.org/10.1038/s41598-022-14382-9>.
- Hague MTJ, Mavengere H, Matute DR, Cooper BS. 2020. Environmental and genetic contributions to imperfect wMel-like *Wolbachia* transmission and frequency variation. *Genetics.* 215:1117–1132. <https://doi.org/10.1534/genetics.120.303330>.
- Hague MTJ et al. 2022. Temperature effects on cellular host-microbe interactions explain continent-wide endosymbiont prevalence. *Curr Biol.* 32:878–888.e8. <https://doi.org/10.1016/j.cub.2021.11.065>.
- Hague MTJ, Wheeler TB, Cooper BS. 2024. Comparative analysis of *Wolbachia* maternal transmission and localization in host ovaries. *Commun Biol.* 7:727. <https://doi.org/10.1038/s42003-024-06431-y>.
- Hamm CA et al. 2014. *Wolbachia* do not live by reproductive manipulation alone: infection polymorphism in *Drosophila suzukii* and *D. subpulchrella*. *Mol Ecol.* 23:4871–4885. <https://doi.org/10.1111/mec.12901>.
- Hartman H, Wheeler TB, Cooper BS, Shropshire JD. 2024. Long-term maintenance of diverse *Drosophila* species with and without *Wolbachia* endosymbionts. *protocols.io.* <https://doi.org/10.17504/protocols.io.rm7vzk85rvx1/v1>.
- Haygood R, Turelli M. 2009. Evolution of incompatibility-inducing microbes in subdivided host populations. *Evolution.* 63:432–447. <https://doi.org/10.1111/j.1558-5646.2008.00550.x>.
- Heath BD, Butcher RDJ, Whitfield WGF, Hubbard SF. 1999. Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. *Curr Biol.* 9:313–316. [https://doi.org/10.1016/S0960-9822\(99\)80139-0](https://doi.org/10.1016/S0960-9822(99)80139-0).
- Heath T, Hedtke SM, Hillis D. 2008. Taxon sampling and the accuracy of phylogenetic analyses. *J Syst Evol.* 46:239–257. <https://doi.org/10.3724/SP.J.1002.2008.08016>.
- Hertig M, Wolbach SB. 1924. Studies on Rickettsia-like microorganisms in insects. *J Med Res.* 44:329–374.
- Hien NT et al. 2021. Environmental factors influence the local establishment of *Wolbachia* in *Aedes aegypti* mosquitoes in two small communities in central Vietnam. *Gates Open Res.* 5:147. <https://doi.org/10.12688/gatesopenres.13347.2>.
- Ho SYW, Phillips MJ, Cooper A, Drummond AJ. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol.* 22:1561–1568. <https://doi.org/10.1093/molbev/msi145>.
- Hodges JL, Lehmann EL. 1963. Estimates of location based on rank tests. *Ann Math Stat.* 34:598–611. <https://doi.org/10.1214/aoms/1177704172>.
- Hoerauf A et al. 2000. Targeting of *Wolbachia* endobacteria in *Litomosoides sigmodontis*: comparison of tetracyclines with chloramphenicol, macrolides and ciprofloxacin. *Trop Med Int Health.* 5:275–279. <https://doi.org/10.1046/j.1365-3156.2000.00544.x>.
- Hoffmann AA. 1988. Partial cytoplasmic incompatibility between two Australian populations of *Drosophila melanogaster*. *Entomol Exp Appl.* 48:61–67. <https://doi.org/10.1111/j.1570-7458.1988.tb02299.x>.
- Hoffmann AA, Cooper BS. 2024. Describing endosymbiont-host interactions within the parasitism-mutualism continuum. *Ecol Evol.* 14:e11705. <https://doi.org/10.1002/ece3.11705>.
- Hoffmann AA, Cooper BS. 2025. Changes in the frequency of facultative endosymbionts in insect populations: overview and applications. *Entomol Gen.* 45:351–368. <https://doi.org/10.1127/entomologia/2025/3052>.
- Hoffmann AA et al. 2011. Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. *Nature.* 476:454–457. <https://doi.org/10.1038/nature10356>.
- Hoffmann AA, Turelli M, Harshman L. 1990. Factors affecting the distribution of cytoplasmic incompatibility in *Drosophila simulans*. *Genetics.* 126:933–948. <https://doi.org/10.1093/genetics/126.4.933>.
- Hoffmann AA, Turelli M, Simmons G. 1986. Unidirectional incompatibility between populations of *Drosophila simulans*. *Evolution.* 40:692–701. <https://doi.org/10.2307/2408456>.
- Hohna S, Landis MJ, Heath TA. 2017. Phylogenetic inference using RevBayes. *Curr Protoc Bioinformatics.* 57:6.16.1–6.16.34. <https://doi.org/10.1002/cpbi.22>.
- Hohna S et al. 2016. RevBayes: bayesian phylogenetic inference using graphical models and an interactive model-specification

- language. *Syst Biol.* 65:726–736. <https://doi.org/10.1093/sysbio/syw021>.
- Holder T. 2023. Pymol ScRipt Collection (PSICO). <https://github.com/speleo3/pymol-psico>
- Hollander M, Wolfe DA, Chicken E. 2015. Nonparametric statistical methods. 1st ed. Wiley. (Wiley Series in Probability and Statistics). <https://doi.org/10.1002/9781119196037>.
- Hudson RR, Coyne JA. 2002. Mathematical consequences of the genealogical species concept. *Evolution.* 56:1557–1565. <https://doi.org/10.1111/j.0014-3820.2002.tb01467.x>.
- Huigens ME, de Almeida RP, Boons PAH, Luck RF, Stouthamer R. 2004. Natural interspecific and intraspecific horizontal transfer of parthenogenesis-inducing *Wolbachia* in *Trichogramma* wasps. *Proc R Soc Lond B Biol Sci.* 271:509–515. <https://doi.org/10.1098/rspb.2003.2640>.
- Hurst LD, McVean GT. 1996. Clade selection, reversible evolution and the persistence of selfish elements: the evolutionary dynamics of cytoplasmic incompatibility. *Proc R Soc Lond B Biol Sci.* 263: 97–104. <https://doi.org/10.1098/rspb.1996.0016>.
- Hurst TP et al. 2012. Impacts of *Wolbachia* infection on predator prey relationships: evaluating survival and horizontal transfer between wMelPop infected *Aedes aegypti* and its predators. *J Med Entomol.* 49:624–630. <https://doi.org/10.1603/ME11277>.
- Iturbe-Ormaetxe I, Burke GR, Riegler M, O'Neill SL. 2005. Distribution, expression, and motif variability of ankyrin domain genes in *Wolbachia pipientis*. *J Bacteriol.* 187:5136–5145. <https://doi.org/10.1128/JB.187.15.5136-5145.2005>.
- Jackman SD et al. 2017. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res.* 27:768–777. <https://doi.org/10.1101/gr.214346.116>.
- Jaenike J, Polak M, Fiskin A, Helou M, Minhas M. 2007. Interspecific transmission of endosymbiotic *Spiroplasma* by mites. *Biol Lett.* 3: 23–25. <https://doi.org/10.1098/rsbl.2006.0577>.
- Jiggins FM, von der Schulenburg JHG, Hurst GDD, Majerus MEN. 2001. Recombination confounds interpretations of *Wolbachia* evolution. *Proc R Soc Lond B Biol Sci.* 268:1423–1427. <https://doi.org/10.1098/rspb.2001.1656>.
- Joshi NA, Fass JN. 2011. Sickie: a sliding-window, adaptive, quality-based trimming tool for FastQ files.
- Jumper J et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature.* 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
- Kaur R, Leigh BA, Ritchie IT, Bordenstein SR. 2022. The cif proteins from *Wolbachia* prophage WO modify sperm genome integrity to establish cytoplasmic incompatibility. *PLoS Biol.* 20:e3001584. <https://doi.org/10.1371/journal.pbio.3001584>.
- Kaur R, McGarry A, Shropshire JD, Leigh BA, Bordenstein SR. 2024. Prophage proteins alter long noncoding RNA and DNA of developing sperm to induce a paternal-effect lethality. *Science.* 383: 1111–1117. <https://doi.org/10.1126/science.adk9469>.
- Kearse M et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 28:1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>.
- Keightley PD et al. 2009. Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Res.* 19:1195–1201. <https://doi.org/10.1101/gr.091231.109>.
- Kent BN, Funkhouser LJ, Setia S, Bordenstein SR. 2011a. Evolutionary genomics of a temperate bacteriophage in an obligate intracellular bacteria (*Wolbachia*). *PLoS One.* 6:e24984. <https://doi.org/10.1371/journal.pone.0024984>.
- Kent BN et al. 2011b. Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*) determined by targeted genome capture. *Genome Biol Evol.* 3:209–218. <https://doi.org/10.1093/gbe/evr007>.
- Kittayapong P, Jamnongluk W, Thipaksorn A, Milne JR, Sindhusake C. 2003. *Wolbachia* infection complexity among insects in the tropical rice-field community. *Mol Ecol.* 12:1049–1060. <https://doi.org/10.1046/j.1365-294X.2003.01793.x>.
- Klasson L et al. 2008. Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. *Mol Biol Evol.* 25:1877–1887. <https://doi.org/10.1093/molbev/msn133>.
- Klasson L et al. 2009. The mosaic genome structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*. *Proc Natl Acad Sci U S A.* 106:5725–5730. <https://doi.org/10.1073/pnas.0810753106>.
- Knizewski L, Kinch LN, Grishin NV, 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>.
- Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW. 2006. GARD: a genetic algorithm for recombination detection. *Bioinformatics.* 22:3096–3098. <https://doi.org/10.1093/bioinformatics/btl474>.
- Kriesner P, Hoffmann AA, Lee SF, Turelli M, Weeks AR. 2013. Rapid sequential spread of two *Wolbachia* variants in *Drosophila simulans*. *PLoS Pathog.* 9:e1003607. <https://doi.org/10.1371/journal.ppat.1003607>.
- Kryazhimskiy S, Plotkin JB. 2008. The population genetics of dN/dS. *PLoS Genet.* 4:e1000304. <https://doi.org/10.1371/journal.pgen.1000304>.
- Kunsch HR. 1989. The jackknife and the bootstrap for general stationary observations. *Ann Stat.* 17:1217–1241. <https://doi.org/10.1214/aos/1176347265>.
- Lachaise D et al. 2000. Evolutionary novelties in islands: *Drosophila santomea*, a new *melanogaster* sister species from Sao Tome. *Proc R Soc Lond B Biol Sci.* 267:1487–1495. <https://doi.org/10.1098/rspb.2000.1169>.
- Lau YL et al. 2015. Draft genome of *Brugia pahangi*: high similarity between *B. pahangi* and *B. malayi*. *Parasit Vectors.* 8:451. <https://doi.org/10.1186/s13071-015-1064-2>.
- Lefoulon E et al. 2016. Breakdown of coevolution between symbiotic bacteria *Wolbachia* and their filarial hosts. *PeerJ.* 4:e1840. <https://doi.org/10.7717/peerj.1840>.
- Lehmann EL. 1975. Nonparametrics: statistical methods based on ranks. Holden-Day. (Holden-Day series in probability and statistics).
- Lenharo M. 2023. Dengue rates drop after release of modified mosquitoes in Colombia. *Nature.* 623:235–236. <https://doi.org/10.1038/d41586-023-03346-2>.
- LePage DP et al. 2017. Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. *Nature.* 543: 243–247. <https://doi.org/10.1038/nature21391>.
- Lewontin RC. 1970. The units of selection. *Annu Rev Ecol Syst.* 1:1–18. <https://doi.org/10.1146/annurev.es.01.110170.000245>.
- Li SJ et al. 2017. Plant-mediated horizontal transmission of *Wolbachia* between whiteflies. *ISME J.* 11:1019–1028. <https://doi.org/10.1038/ismej.2016.164>.
- Liang H, Zhou W, Landweber LF. 2006. SWAKK: a web server for detecting positive selection in proteins using a sliding window substitution rate analysis. *Nucleic Acids Res.* 34:W382–W384. <https://doi.org/10.1093/nar/gkl272>.
- Lindsey ARI et al. 2018. Evolutionary genetics of cytoplasmic incompatibility genes cifA and cifB in prophage WO of *Wolbachia*.

- Genome Biol Evol. 10:434–451. <https://doi.org/10.1093/gbe/evy012>.
- Liu RY, Singh K. 1992. Moving blocks jackknife and bootstrap capture weak dependence. In: LePage R, Billard L, editors. Exploring the limits of bootstrap. Wiley. p. 225–248. <https://cir.nii.ac.jp/crid/1573668924689635584>
- Loytynoja A. 2014. Phylogeny-aware alignment with PRANK. Methods Mol Biol. 1079:155–170. https://doi.org/10.1007/978-1-62703-646-7_10.
- Maddison WP. 1997. Gene trees in species trees. Syst Biol. 46:523–536. <https://doi.org/10.1093/sysbio/46.3.523>.
- Madhav M, Parry R, Morgan JAT, James P, Asgari S. 2020. *Wolbachia* endosymbiont of the horn fly (*Haematobia irritans irritans*): a super-group A strain with multiple horizontally acquired cytoplasmic incompatibility genes. Appl Environ Microbiol. 86:e02589-19. <https://doi.org/10.1128/AEM.02589-19>.
- Manoj RRS, Latrofa MS, Epis S, Otranto D. 2021. *Wolbachia*: endosymbiont of onchocercid nematodes and their vectors. Parasit Vectors. 14:245. <https://doi.org/10.1186/s13071-021-04742-1>.
- Martinez J et al. 2022. Genome sequencing and comparative analysis of *Wolbachia* strain wAlbA reveals *Wolbachia*-associated plasmids are common. PLoS Genet. 18:e1010406. <https://doi.org/10.1371/journal.pgen.1010406>.
- Martinez J, Klasson L, Welch JJ, Jiggins FM. 2021. Life and death of selfish genes: comparative genomics reveals the dynamic evolution of cytoplasmic incompatibility. Mol Biol Evol. 38:2–15. <https://doi.org/10.1093/molbev/msaa209>.
- Martinez J et al. 2014. Symbionts commonly provide broad spectrum resistance to viruses in insects: a comparative analysis of *Wolbachia* strains. PLoS Pathog. 10:e1004369. <https://doi.org/10.1371/journal.ppat.1004369>.
- Masui S, Kamoda S, Sasaki T, Ishikawa H. 2000. Distribution and evolution of bacteriophage WO in *Wolbachia*, the endosymbiont causing sexual alterations in arthropods. J Mol Evol. 51:491–497. <https://doi.org/10.1007/s002390010112>.
- Matute DR, Cooper BS. 2021. Comparative studies on speciation: 30 years since Coyne and Orr. Evolution. 75:764–778. <https://doi.org/10.1111/evo.14181>.
- Meany MK et al. 2019. Loss of cytoplasmic incompatibility and minimal fecundity effects explain relatively low *Wolbachia* frequencies in *Drosophila mauritiana*. Evolution. 73:1278–1295. <https://doi.org/10.1111/evo.13745>.
- Miller WJ, Riegler M. 2006. Evolutionary dynamics of wAu-like *Wolbachia* variants in neotropical *Drosophila* spp. Appl Environ Microbiol. 72:826–835. <https://doi.org/10.1128/AEM.72.1.826-835.2006>.
- Moledo Gesto JS et al. 2021. Large-scale deployment and establishment of *Wolbachia* into the *Aedes aegypti* population in Rio de Janeiro, Brazil. Front Microbiol. 12:711107. <https://doi.org/10.3389/fmicb.2021.711107>.
- Moran NA. 2007. Symbiosis as an adaptive process and source of phenotypic complexity. Proc Natl Acad Sci U S A. 104:8627–8633. <https://doi.org/10.1073/pnas.0611659104>.
- Nice CC, Gompert Z, Forister ML, Fordyce JA. 2009. An unseen foe in arthropod conservation efforts: the case of *Wolbachia* infections in the Karner blue butterfly. Biol Conserv. 142:3137–3146. <https://doi.org/10.1016/j.biocon.2009.08.020>.
- Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics. 148:929–936. <https://doi.org/10.1093/genetics/148.3.929>.
- Ochman H, Elwyn S, Moran NA. 1999. Calibrating bacterial evolution. Proc Natl Acad Sci U S A. 96:12638–12643. <https://doi.org/10.1073/pnas.96.22.12638>.
- Ochman H, Wilson AC. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J Mol Evol. 26:74–86. <https://doi.org/10.1007/BF02111283>.
- Oladipupo SO, Hochstrasser M. 2025. Deubiquitylases and nucleases in bacterial symbiont-induced cytoplasmic incompatibility. Biochem Soc Trans. 53:1365–1375. <https://doi.org/10.1042/BST20253047>.
- Oliveira DCSG, Raychoudhury R, Lavrov DV, Werren JH. 2008. Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp *Nasonia* (Hymenoptera: Pteromalidae). Mol Biol Evol. 25:2167–2180. <https://doi.org/10.1093/molbev/msn159>.
- Oliver KM, Degnan PH, Burke GR, Moran NA. 2010. Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. Annu Rev Entomol. 55:247–266. <https://doi.org/10.1146/annurev-ento-112408-085305>.
- O'Neill SL, Giordano R, Colbert AM, Karr TL, Robertson HM. 1992. 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. Proc Natl Acad Sci U S A. 89:2699–2702. <https://doi.org/10.1073/pnas.89.7.2699>.
- Oswald K, Oladipupo S, Mendez L, Gillespie JJ, Beckmann JF. 2025 Mar 10. Origins of cin: Lateral gene transfer of cytoplasmic incompatibility nuclease operon to *Orientia tsutsugamushi*. bioRxiv 641471. <https://doi.org/10.1101/2025.03.04.641471>, preprint; not peer reviewed.
- Porter J, Sullivan W. 2023. The cellular lives of *Wolbachia*. Nat Rev Microbiol. 21:750–766. <https://doi.org/10.1038/s41579-023-00918-x>.
- Prout T. 1994. Some evolutionary possibilities for a microbe that causes incompatibility in its host. Evolution. 48:909–911. <https://doi.org/10.2307/2410496>.
- Qing X et al. 2025. Phylogenomic insights into the evolution and origin of Nematoda. Syst Biol. 74:349–358. <https://doi.org/10.1093/sysbio/syae073>.
- Radousky YA et al. 2023. Distinct *Wolbachia* localization patterns in oocytes of diverse host species reveal multiple strategies of maternal transmission. Genetics. 224:iyad038. <https://doi.org/10.1093/genetics/iyad038>.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol. 67:901–904. <https://doi.org/10.1093/sysbio/syy032>.
- Rannala B, Huelsenbeck JP, Yang Z, Nielsen R. 1998. Taxon sampling and the accuracy of large phylogenies. Syst Biol. 47:702–710. <https://doi.org/10.1080/106351598260680>.
- Raychoudhury R, Baldo L, Oliveira DCSG, Werren JH. 2009. Modes of acquisition of *Wolbachia*: horizontal transfer, hybrid introgression, and codivergence in the *Nasonia* species complex. Evolution. 63:165–183. <https://doi.org/10.1111/j.1558-5646.2008.00533.x>.
- Razavi M, Kristiansson E, Flach C-F, Larsson DGJ. 2020. The association between insertion sequences and antibiotic resistance genes. mSphere. 5:e00418-20. <https://doi.org/10.1128/mSphere.00418-20>.
- Reveillaud J et al. 2019. The *Wolbachia* mobilome in *Culex pipiens* includes a putative plasmid. Nat Commun. 10:1051. <https://doi.org/10.1038/s41467-019-08973-w>.
- Reynolds KT, Hoffmann AA. 2002. 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. <https://doi.org/10.1017/S0016672302005827>.
- Richardson MF et al. 2012. Population genomics of the *Wolbachia* endosymbiont in *Drosophila melanogaster*. PLoS Genet. 8:e1003129. <https://doi.org/10.1371/journal.pgen.1003129>.

- Riegler M, Sidhu M, Miller WJ, O'Neill SL. 2005. Evidence for a global *Wolbachia* replacement in *Drosophila melanogaster*. *Curr Biol*. 15: 1428–1433. <https://doi.org/10.1016/j.cub.2005.06.069>.
- Ronquist F et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 61:539–542. <https://doi.org/10.1093/sysbio/sys029>.
- Ross PA, Ritchie SA, Axford JK, Hoffmann AA. 2019. Loss of cytoplasmic incompatibility in *Wolbachia*-infected *Aedes aegypti* under field conditions. *PLoS Negl Trop Dis*. 13:e0007357. <https://doi.org/10.1371/journal.pntd.0007357>.
- Ross PA et al. 2017. *Wolbachia* infections in *Aedes aegypti* differ markedly in their response to cyclical heat stress. *PLOS Pathog*. 13: e1006006. <https://doi.org/10.1371/journal.ppat.1006006>.
- Russell JA et al. 2009. Specialization and geographic isolation among *Wolbachia* symbionts from ants and lycaenid butterflies. *Evolution*. 63:624–640. <https://doi.org/10.1111/j.1558-5646.2008.00579.x>.
- Scholz M et al. 2020. Large scale genome reconstructions illuminate *Wolbachia* evolution. *Nat Commun*. 11:5235. <https://doi.org/10.1038/s41467-020-19016-0>.
- Schrodinger LLC. 2015. The PyMOL Molecular Graphics System, Version 2.5.2. <https://www.pymol.org>
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Shi P-Q et al. 2024. Rickettsia transmission from whitefly to plants benefits herbivore insects but is detrimental to fungal and viral pathogens. *mBio*. 15:e0244823. <https://doi.org/10.1128/mbio.02448-23>.
- Shropshire JD. 2025. Preparing fly-food spoons for *Drosophila* embryo collection. *protocols.io*. <https://doi.org/10.17504/protocols.io.bp2l6ywerqvq/v1>
- Shropshire JD, Bordenstein SR. 2019. Two-By-One model of cytoplasmic incompatibility: synthetic recapitulation by transgenic expression of *cifA* and *cifB* in *Drosophila*. *PLOS Genet*. 15:e1008221. <https://doi.org/10.1371/journal.pgen.1008221>.
- Shropshire JD, Hamant E, Conner WR, Cooper BS. 2022. *cifB*-transcript levels largely explain cytoplasmic incompatibility variation across divergent *Wolbachia*. *PNAS Nexus*. 1:pgac099. <https://doi.org/10.1093/pnasnexus/pgac099>.
- Shropshire JD, Hamant E, Cooper BS. 2021b. Male age and *Wolbachia* dynamics: investigating how fast and why bacterial densities and cytoplasmic incompatibility strengths vary. *mBio*. 12:e02998-21. <https://doi.org/10.1128/mBio.02998-21>.
- Shropshire JD, Kalra M, Bordenstein SR. 2020b. Evolution-guided mutagenesis of the cytoplasmic incompatibility proteins: identifying CifA's complex functional repertoire and new essential regions in CifB. *PLOS Pathog*. 16:e1008794. <https://doi.org/10.1371/journal.ppat.1008794>.
- Shropshire JD, Leigh B, Bordenstein SR. 2020a. Symbiont-mediated cytoplasmic incompatibility: what have we learned in 50 years? *eLife*. 9:e61989. <https://doi.org/10.7554/eLife.61989>.
- Shropshire JD et al. 2019. Models and nomenclature for cytoplasmic incompatibility: caution over premature conclusions—a response to Beckmann et al. *Trends Genet*. 35:397–399. <https://doi.org/10.1016/j.tig.2019.03.004>.
- Shropshire JD, On J, Layton EM, Zhou H, Bordenstein SR. 2018. One prophage WO gene rescues cytoplasmic incompatibility in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 115:4987–4991. <https://doi.org/10.1073/pnas.1800650115>.
- Shropshire JD, Rosenberg R, Bordenstein SR. 2021a. The impacts of cytoplasmic incompatibility factor (*cifA* and *cifB*) genetic variation on phenotypes. *Genetics*. 217:1–13. <https://doi.org/10.1093/genetics/iyaa007>.
- Siguier P, Goubeyre E, Chandler M. 2014. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev*. 38:865–891. <https://doi.org/10.1111/1574-6976.12067>.
- Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Sintupachee S, Milne JR, Poonchaisri S, Baimai V, Kittayapong P. 2006. Closely related *Wolbachia* strains within the pumpkin arthropod community and the potential for horizontal transmission via the plant. *Microb Ecol*. 51:294–301. <https://doi.org/10.1007/s00248-006-9036-x>.
- Small ST et al. 2019. Human migration and the spread of the nematode parasite *Wuchereria bancrofti*. *Mol Biol Evol*. 36:1931–1941. <https://doi.org/10.1093/molbev/msz116>.
- Sokal A. 1997. Monte Carlo methods in statistical mechanics: foundations and new algorithms. In: DeWitt-Morette C, Cartier P, Folacci A, editors. *Functional integration: basics and applications*. Springer US. p. 131–192. https://doi.org/10.1007/978-1-4899-0319-8_6.
- Strunov A, Schmidt K, Kapun M, Miller WJ. 2022. Autophagy regulates endosymbiont distribution in early *Drosophila* embryogenesis. *Autophagy Rep*. 1:373–376. <https://doi.org/10.1080/27694127.2022.2112124>.
- Sun G, Zhang M, Chen H, Hochstrasser M. 2022. The CinB nuclease from *wNo Wolbachia* is sufficient for induction of cytoplasmic incompatibility in *Drosophila*. *mBio*. 13:e03177-21. <https://doi.org/10.1128/mbio.03177-21>.
- Suvorov A et al. 2022. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *Curr Biol*. 32:111–123.e5. <https://doi.org/10.1016/j.cub.2021.10.052>.
- Tamura K, Subramanian S, Kumar S. 2004. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol Biol Evol*. 21:36–44. <https://doi.org/10.1093/molbev/msg236>.
- Tan Y, Aravind L, Zhang D. 2024. Genomic underpinnings of cytoplasmic incompatibility: CIF gene-neighborhood diversification through extensive lateral transfers and recombination in *Wolbachia*. *Genome Biol Evol*. 16:evae171. <https://doi.org/10.1093/gbe/evae171>.
- Taylor MJ, Bandi C, Hoerauf A. 2005. *Wolbachia* bacterial endosymbionts of filarial nematodes. *Adv Parasitol*. 60:245–284. [https://doi.org/10.1016/S0065-308X\(05\)60004-8](https://doi.org/10.1016/S0065-308X(05)60004-8).
- Terretaz K, Horard B, Weill M, Loppin B, Landmann F. 2023. Functional analysis of *Wolbachia* Cid effectors unravels cooperative interactions to target host chromatin during replication. *PLOS Pathog*. 19:e1011211. <https://doi.org/10.1371/journal.ppat.1011211>.
- Thia JA, Dorai APS, Hoffmann AA. 2025. Symbiotic bacteria and pest control: plant toxins, chemical pesticides, and fungal entomopathogens. *Trends Microbiol*. 33:976–988. <https://doi.org/10.1016/j.tim.2025.04.007>.
- Turelli M. 1994. Evolution of incompatibility-inducing microbes and their hosts. *Evolution*. 48:1500–1513. <https://doi.org/10.2307/2410244>.
- Turelli M et al. 2018. Rapid global spread of *wRi*-like *Wolbachia* across multiple *Drosophila*. *Curr Biol*. 28:963–971.e8. <https://doi.org/10.1016/j.cub.2018.02.015>.
- Turelli M, Hoffmann AA. 1991. Rapid spread of an inherited incompatibility factor in California *Drosophila*. *Nature*. 353:440–442. <https://doi.org/10.1038/353440a0>.
- Turelli M, Hoffmann AA. 1995. Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from

- natural populations. *Genetics*. 140:1319–1338. <https://doi.org/10.1093/genetics/140.4.1319>.
- Turelli M, Katznelson A, Ginsberg PS. 2022. Why *Wolbachia*-induced cytoplasmic incompatibility is so common. *Proc Natl Acad Sci U S A*. 119:e2211637119. <https://doi.org/10.1073/pnas.2211637119>.
- Turelli M, Lipkowitz JR, Brandvain Y. 2014. On the Coyne and Orr-igin of species: effects of intrinsic postzygotic isolation, ecological differentiation, X-chromosome size, and sympatry on *Drosophila* speciation. *Evolution*. 68:1176–1187. <https://doi.org/10.1111/evo.12330>.
- Utarini A et al. 2021. Efficacy of *Wolbachia*-infected mosquito deployments for the control of dengue. *N Engl J Med*. 384:2177–2186. <https://doi.org/10.1056/NEJMoa2030243>.
- Van Meer MMM, Stouthamer R. 1999. Cross-order transfer of *Wolbachia* from *Muscidifurax uniraptor* (Hymenoptera: Pteromalidae) to *Drosophila simulans* (Diptera: Drosophilidae). *Heredity* (Edinb). 82: 163–169. <https://doi.org/10.1038/sj.hdy.6884610>.
- Vancaester E, Blaxter M. 2023. Phylogenomic analysis of *Wolbachia* genomes from the Darwin Tree of Life biodiversity genomics project. *PLOS Biol*. 21:e3001972. <https://doi.org/10.1371/journal.pbio.3001972>.
- Vancaester E et al. 2025. Ghosts of symbionts past: the hidden history of the dynamic association between filarial nematodes and their *Wolbachia* endosymbionts. *G3* (Bethesda). 15:jkaf226. <https://doi.org/10.1093/g3journal/jkaf226>.
- Vavre F, Fleury F, Lepetit D, Fouillet P, Bouletreau M. 1999. Phylogenetic evidence for horizontal transmission of *Wolbachia* in host-parasitoid associations. *Mol Biol Evol*. 16:1711–1723. <https://doi.org/10.1093/oxfordjournals.molbev.a026084>.
- Velez ID et al. 2023. Reduced dengue incidence following city-wide *wMel Wolbachia* mosquito releases throughout three Colombian cities: interrupted time series analysis and a prospective case-control study. *PLoS Negl Trop Dis*. 17:e0011713. <https://doi.org/10.1371/journal.pntd.0011713>.
- Venkat A, Hahn MW, Thornton JW. 2018. Multinucleotide mutations cause false inferences of lineage-specific positive selection. *Nat Ecol Evol*. 2:1280–1288. <https://doi.org/10.1038/s41559-018-0584-5>.
- Walker T et al. 2011. The *wMel Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. *Nature*. 476:450–453. <https://doi.org/10.1038/nature10355>.
- Wang X et al. 2020. Phylogenomic analysis of *Wolbachia* strains reveals patterns of genome evolution and recombination. *Genome Biol Evol*. 12:2508–2520. <https://doi.org/10.1093/gbe/evaa219>.
- Wang YH et al. 2016. Fossil record of stem groups employed in evaluating the chronogram of insects (Arthropoda: Hexapoda). *Sci Rep*. 6:38939. <https://doi.org/10.1038/srep38939>.
- Weinert LA, Araujo-Jnr EV, Ahmed MZ, Welch JJ. 2015. The incidence of bacterial endosymbionts in terrestrial arthropods. *Proc R Soc B Biol Sci*. 282:20150249. <https://doi.org/10.1098/rspb.2015.0249>.
- Werren JH, Bartos JD. 2001. Recombination in *Wolbachia*. *Curr Biol*. 11:431–435. [https://doi.org/10.1016/S0960-9822\(01\)00101-4](https://doi.org/10.1016/S0960-9822(01)00101-4).
- Werren JH, Zhang W, Guo L. 1995. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proc R Soc B Biol Sci*. 261:55–63. <https://doi.org/10.1098/rspb.1995.0117>.
- Wheeler TB, Hartman H, Cooper BS, Shropshire JD. 2024. Preparing food for *Drosophila melanogaster* and its relatives. *protocols.io*. <https://doi.org/10.17504/protocols.io.j8nlk8poxl5r/v1>
- Wiegmann BM et al. 2011. Episodic radiations in the fly tree of life. *Proc Natl Acad Sci U S A*. 108:5690–5695. <https://doi.org/10.1073/pnas.1012675108>.
- Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. 2014. Permutation inference for the general linear model. *NeuroImage*. 92:381–397. <https://doi.org/10.1016/j.neuroimage.2014.01.060>.
- Winkler AM, Webster MA, Vidaurre D, Nichols TE, Smith SM. 2015. Multi-level block permutation. *NeuroImage*. 123:253–268. <https://doi.org/10.1016/j.neuroimage.2015.05.092>.
- Wong TKF et al. 2025. IQ-TREE 3: Phylogenomic inference software using complex evolutionary models. *EcoEvoRxiv*. <https://doi.org/10.32942/X2P62N>, preprint; not peer reviewed.
- Wu M et al. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipiens* wMel: a streamlined genome overrun by mobile genetic elements. *PLOS Biol*. 2:E69. <https://doi.org/10.1371/journal.pbio.0020069>.
- Xu J, Zhang Y. 2010. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics*. 26:889–895. <https://doi.org/10.1093/bioinformatics/btq066>.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 13:555–556. <https://doi.org/10.1093/bioinformatics/13.5.555>.
- Yang Z. 2006. Computational molecular evolution. Oxford University Press. (Oxford Series in Ecology and Evolution).
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591. <https://doi.org/10.1093/molbev/msm088>.
- Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol*. 15:496–503. [https://doi.org/10.1016/S0169-5347\(00\)01994-7](https://doi.org/10.1016/S0169-5347(00)01994-7).
- Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*. 17:32–43. <https://doi.org/10.1093/oxfordjournals.molbev.a026236>.
- Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*. 155:431–449. <https://doi.org/10.1093/genetics/155.1.431>.
- Yang Z, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol*. 22:1107–1118. <https://doi.org/10.1093/molbev/msi097>.
- Yen JH, Barr AR. 1971. New hypothesis of cause of cytoplasmic incompatibility in *Culex pipiens* L. *Nature*. 232:657–658. <https://doi.org/10.1038/232657a0>.
- Yen JH, Barr AR. 1973. The etiological agent of cytoplasmic incompatibility in *Culex pipiens*. *J Invertebr Pathol*. 22:242–250. [https://doi.org/10.1016/0022-2011\(73\)90141-9](https://doi.org/10.1016/0022-2011(73)90141-9).
- Zhang C, Mirarab S. 2022. Weighting by gene tree uncertainty improves accuracy of quartet-based species trees. *Mol Biol Evol*. 39:msac215. <https://doi.org/10.1093/molbev/msac215>.
- Zhang Y, Skolnick J. 2004. Scoring function for automated assessment of protein structure template quality. *Proteins*. 57: 702–710. <https://doi.org/10.1002/prot.20264>.
- Zimmermann L et al. 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol*. 430:2237–2243. <https://doi.org/10.1016/j.jmb.2017.12.007>.