Loss of cytoplasmic incompatibility and minimal fecundity effects explain relatively low *Wolbachia* frequencies in *Drosophila mauritiana*

Megan K. Meany[1], William R. Conner[1], Sophia V. Richter[1], Jessica A. Bailey[1], Michael Turelli[2] and Brandon S. Cooper[1¶]

[1]Division of Biological Sciences, University of Montana, Missoula, MT USA

[2]Department of Evolution and Ecology, University of California, Davis, CA USA

Correspondence:

[¶]Division of Biological Sciences

University of Montana

32 Campus Drive, ISB

Missoula, MT 59812, USA

Email: brandon.cooper@umontana.edu

RUNNING TITLE: *w*Mau *Wolbachia* in *Drosophila mauritiana*

KEYWORDS

host-microbe interactions, introgression, spatial spread, *w*Mau *Wolbachia*, WO phage

ABSTRACT

Maternally transmitted *Wolbachia* bacteria infect about half of all insect species. Many *Wolbachia* cause cytoplasmic incompatibility (CI), reduced egg hatch when uninfected females mate with infected males. CI produces a frequency-dependent fitness advantage that leads to high equilibrium *Wolbachia* frequencies but does not aid *Wolbachia* spread from low frequencies. The fitness advantages that produce initial *Wolbachia* spread and maintain non-CI *Wolbachia* remain elusive. *Wolbachia* variant *w*Mau that infects *Drosophila mauritiana* does not cause CI, despite being very closely related to the CI-causing *Wolbachia w*No in *D. simulans* (0.068% sequence divergence over 682,494 bp), suggesting that CI was recently lost. Using draft *Wolbachia* genomes, we identify a deletion in a WO phage gene that may disrupt CI. We find no evidence that *w*Mau increases host fecundity. We report intermediate and apparently stable *w*Mau infection frequencies on the Indian Ocean island of Mauritius, consistent with no CI and no appreciable positive effect on fecundity. Our data indicate that *w*Mau frequencies reflect a balance between unknown, weak positive effects on fitness and imperfect maternal transmission. Phylogenomic analyses suggest that Supergroup B *Wolbachia*, including *w*Mau, diverged from Supergroup A *Wolbachia*, including *w*Mel from *D. melanogaster*, 6–46 million years ago, later than previously estimated.

INTRODUCTION

Maternally transmitted *Wolbachia* infect about half of all insect species (Werren and Windsor 2000; Zug and Hammerstein 2012; Weinert et al. 2015), as well as other arthropods (Jeyaprakash and Hoy 2000; Hilgenboecker et al. 2008) and nematodes (Taylor et al. 2013). Host species may acquire *Wolbachia* from common ancestors, from sister species via hybridization and introgression, or horizontally (O'Neill et al. 1992; Rousset and Solignac 1995; Huigens et al. 2004; Baldo et al. 2008; Raychoudhury et al. 2009; Gerth and Bleidorn 2016; Schuler et al. 2016; Turelli et al. 2018). Many *Wolbachia* cause cytoplasmic incompatibility (CI), reduced egg hatch when uninfected females mate with *Wolbachia*-infected males. Three parameters usefully approximate the frequency dynamics and equilibria of CI-causing *Wolbachia*: the relative hatch rate of uninfected eggs fertilized by infected males (H), the fitness of infected females relative to uninfected females (F), and the proportion of uninfected ova produced by infected females (μ) (Caspari and Watson 1959; Hoffmann et al. 1990). To spread deterministically from low frequencies, *Wolbachia* must produce $F(1-\mu) > 1$, irrespective of CI. Once they become sufficiently common, CI-causing infections, such as *w*Ri-like *Wolbachia* in *Drosophila simulans* and several other *Drosophila* (Turelli et al. 2018), spread to high equilibrium frequencies, dominated by a balance between CI and imperfect maternal transmission (Turelli and Hoffmann 1995; Kreisner et al. 2016). In contrast, non-CI-causing *Wolbachia*, such as *w*Au in *D. simulans*, typically persist at lower equilibrium frequencies, maintained by a balance between positive *Wolbachia* effects on host fitness and imperfect maternal transmission (Hoffmann and Turelli 1997; Kreisner et al. 2013). When $H < F(1-\mu) < 1$, "bistable" dynamics result, producing stable equilibria at 0 and at a higher frequency denoted p_s, where $0.50 < p_s \leq 1$ (Turelli and Hoffmann 1995). Bistability explains the pattern and rate of spread of *w*Mel transinfected into *Aedes aegypti* to suppress the spread of dengue, Zika and other human diseases (Hoffmann et al. 2011; Barton and Turelli 2011; Turelli and Barton 2017; Schmidt et al. 2017).

In contrast to the bistability observed with *w*Mel transinfections, natural *Wolbachia* infections seem to spread via "Fisherian" dynamics with $F(1-\mu) > 1$ (Fisher 1937; Kriesner et al. 2013; Hamm et al. 2014). Several *Wolbachia* effects could generate $F(1-\mu) > 1$, but we do not yet know which ones actually do. For example, *w*Ri has evolved to increase *D. simulans* fecundity in only a few decades (Weeks et al. 2007), *w*Mel seems to enhance *D. melanogaster* fitness in high and low iron environments (Brownlie et al. 2009), and several *Wolbachia*

including *w*Mel protect their *Drosophila* hosts from RNA viruses (Hedges et al. 2008; Teixeira et al. 2008; Martinez et al. 2014). However, it remains unknown which if any these potential fitness benefits underlie *Wolbachia* spread in nature. For instance, *w*Mel seems to have little effect on viral abundance in wild-caught *D. melanogaster* (Shi et al. 2018).

D. mauritiana, D. simulans and *D. sechellia* comprise the *D. simulans* clade within the nine-species *D. melanogaster* subgroup of *Drosophila*. The *D. simulans* clade diverged from *D. melanogaster* approximately three million years ago (mya), with the island endemics *D. sechellia* (Seychelles archipelago) and *D. mauritiana* (Mauritius) thought to originate in only the last few hundred thousand years (Lachaise et al. 1986; Ballard 2000a; Dean and Ballard 2004; McDermott and Kliman 2008; Garrigan et al. 2012; Brand et al. 2013; Garrigan et al. 2014). *D. simulans* is widely distributed around the globe, but has never been collected on Mauritius (David et al. 1989; Legrand et al. 2011). However, evidence of mitochondrial and nuclear introgression supports interisland migration and hybridization between these species (Ballard 2000a; Nunes et al. 2010; Garrigan et al. 2012), which could allow introgressive *Wolbachia* transfer (Rousset and Solignac 1995).

D. mauritiana is infected with *Wolbachia* denoted *w*Mau, likely acquired via introgression from other *D. simulans*-clade hosts (Rousset and Solignac 1995). *Wolbachia* variant *w*Mau may also infect *D. simulans* in Madagascar and elsewhere in Africa and the South Pacific (denoted *w*Ma in *D. simulans*) (Ballard 2000a; Ballard 2004). *w*Mau does not cause CI in *D. mauritiana* or when transinfected into *D. simulans* (Giordano et al. 1995), despite being very closely related to *w*No strains that do cause CI in *D. simulans* (Merçot et al. 1995; Rousset and Solignac 1995; James and Ballard 2000). (Also, *D. simulans* seems to be a "permissive" host for CI, as evidenced by the fact that *w*Mel, which causes little CI in its native host, *D. melanogaster*, causes intense CI in *D. simulans* [Poinsot et al. 1998].) In contrast to the lack of CI, Fast et al. (2011) reported that a *w*Mau variant increased *D. mauritiana* fecundity four-fold. This fecundity effect occurred in concert with *w*Mau-induced alternations of programmed cell death in the germarium and of germline stem cell mitosis, providing insight into the mechanisms underlying increased egg production (Fast et al. 2011). However, the generality of these effects across *w*Mau variants, and whether such extreme effects are plausible mechanisms of *w*Mau spread, remains unknown.

Here, we assess the genetic and phenotypic basis of *w*Mau frequencies in *D. mauritiana* on Mauritius by combining analysis of *w*Mau draft genomes with analysis of *w*Mau effects on host fecundity and egg hatch. We identify a single mutation that we predict disrupts CI in *w*Mau, relative to its apparent sister *w*No. The loss of CI in the *w*Mau lineage is consistent with theory demonstrating that selection within host species does not act to increase or maintain the level of CI (Prout 1994; Turelli 1994; Haygood and Turelli 2009), but instead acts to increase $F(1 - \mu)$, the product of *Wolbachia* effects on host fitness and maternal transmission efficiency (Turelli 1994). The loss of CI helps explain the intermediate *w*Mau frequencies on Mauritius, reported by us and Giordano et al. (1995). We find no *w*Mau effects on host fecundity, suggesting that the results of Fast et al. (2011) may be anomalous. Indeed, theoretical analyses show that even a two-fold fecundity increase cannot be reconciled with the observed intermediate population frequencies, unless maternal *w*Mau transmission is exceptionally unreliable. Finally, we report the unexpected maintenance of two distinct classes of mtDNA haplotypes among *Wolbachia*-uninfected *D. mauritiana* and present theoretical analyses illustrating why this is unexpected.

MATERIALS AND METHODS

Drosophila Husbandry and Stocks

The *D. mauritiana* isofemale lines used in this study ($N = 32$) were sampled from Mauritius in 2006 by Margarita Womack and kindly provided to us by Prof. Daniel Matute at the University of North Carolina in Chapel Hill. Stocks were maintained on modified version of the standard Bloomington-cornmeal medium (Bloomington Stock Center, Bloomington, IN) and were kept at 25°C, 12 light:12 dark photoperiod prior to the start of our experiments.

Determining *Wolbachia* infection status and comparing infection frequencies

DNA was extracted from each isofemale line using a standard 'squish' buffer protocol (Gloor et al. 1993), and infection status was determined using a polymerase chain reaction (PCR) assay (Simpliamp ThermoCycler, Applied Biosystems, Singapore). We used the GoTaq™ DNA Polymerase (Promega™, Wisconsin, USA) master mix, and PCR began with 3 minutes at 94°C, followed by 34 rounds of 30 seconds at 94°C, 30 seconds at 55°C, and 1 minute and 15 seconds at 72°C. The profile finished with 8 minutes at 72°C. We amplified the *Wolbachia*-specific *wsp* gene and a nuclear control region of the *2L* chromosome; primers are listed in Supplementary

Table 1. PCR products were visualized using 1% agarose gels that included a molecular-weight ladder. Assuming a binomial distribution, we estimated exact 95% binomial confidence intervals for the infection frequencies on Mauritius. Using Fisher's Exact Test, we tested for temporal variation in *w*Mau frequencies by comparing our frequency estimate to a previous estimate (Giordano et al. 1995). All analyses were performed using R version 3.5.1 (R Team 2015).

We used quantitative PCR (qPCR) (MX3000P, Agilent Technologies, Germany) to confirm that tetracycline-treated flies were cleared of *w*Mau. DNA was extracted from *D. mauritiana* flies after four generations of tetracycline treatment, as described below. Our qPCR used a PowerUp™ SYBR™ Green Master Mix (Applied Biosystems™, California, USA). qPCR began with 2 minutes at 50℃ and 10 minutes at 95℃, followed by 40 cycles of 10 seconds at 95℃ and 30 seconds at 60℃. We finished with 15 seconds at 95℃, 15 seconds at 55℃, and 15 seconds at 95℃. Primers are listed in Supplementary Table 1.

Wolbachia DNA extraction, library preparation, and sequencing

We sequenced *w*Mau-infected *R9*, *R29*, and *R60 D. mauritiana* genotypes. Tissue samples for genomic DNA were extracted using a modified CTAB Genomic DNA Extraction protocol. Pools of flies ($N = 5$), including males and females, were homogenized in 200 ml CTAB solution (100mM Tris HCl pH 8.0, 10nM EDTA, 1.4M NaCl, CTAB 2% (w/v)) using motorized pestle. We incubated each sample for 10 min at 65°C, after which we added 200 µl of chloroform; samples were then centrifuged at 13,000 rpm for 5 min at 4°C. We transferred the aqueous phase into a new tube, added 1:1 volume isopropanol, incubated samples on ice for 1 hr, and centrifuged them at 13,000 rpm for 30 min at 4°C. We discarded the supernatant, added 500 µl 70% EtOH, and then inverted the samples 2–3 times to re-suspend the pellets. Finally, we centrifuged the samples at 13,000 rpm for 15 min at 4°C, incubated at room temperature for 15 min, and centrifuged again for 5 min. We discarded the ethanol and repeated these final steps. The resulting pellet was air dried at room temperature for 5 min and dissolved in 40ul 1× TE Buffer.

DNA quantity was tested on an Implen Nanodrop (Implen, München, Germany) and total DNA was quantified by Qubit Fluorometric Quantitation (Invitrogen, Carlsbad, California, USA). DNA was cleaned using Agencourt AMPure XP beads (Beckman Coulter, Inc., Brea, CA, U.S.A), following manufacturers' instructions, and eluted in 50 µl 1× TE Buffer for shearing.

DNA was sheared using a Covaris E220 Focused Ultrasonicator (Covaris Inc., Woburn, MA) to a target size of 400 bp.

We prepared libraries using NEBNext® Ultra™ II DNA Library Prep with Sample Purification Beads (New England BioLabs, Ipswich, Massachusetts). End Preparation, Adaptor Ligation, Bead Clean-up with size selection, PCR Enrichment of Adaptor-Ligated DNA, and Clean-up of PCR Reaction were completed following the manufacturer's protocols. Final fragment sizes and concentrations were confirmed using a TapeStation 2200 system (Agilent, Santa Clara, California). We indexed samples using NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 3 & Index Primers Set 4), and 10 µl of each sample was shipped to Novogene (Sacramento, CA) for sequencing using Illumina HiSeq 4000 (San Diego, CA), generating paired-end, 150 bp reads.

Wolbachia assembly

We obtained published reads ($N = 6$) from Garrigan et al. (2014), and assembled these genomes along with the *R9*, *R29*, and *R60* genomes that we sequenced. Reads were trimmed using Sickle v. 1.33 (Joshi and Fass 2011) and assembled using ABySS v. 2.0.2 (Jackman et al. 2017). *K* values of 41, 51, 61, and 71 were used, and scaffolds with the best nucleotide BLAST matches to known *Wolbachia* sequences with E-values less than 10^{-10} were extracted as the draft *Wolbachia* assemblies. We deemed samples infected if the largest *Wolbachia* assembly was at least 1 million bases and uninfected if the largest assembly was fewer than 100,000 bases. No samples produced *Wolbachia* assemblies between 100,000 and 1 million bases. Of the six sets of published reads we analyzed (Garrigan et al. 2014), only lines *R31* and *R41* were *w*Mau-infected.

To assess the quality of our draft assemblies, we used BUSCO v. 3.0.0 to search for orthologs of the near-universal, single-copy genes in the BUSCO proteobacteria database (Simao et al. 2015). As a control, we performed the same search using the reference genomes for *w*Ri (Klasson et al. 2009), *w*Au (Sutton et al. 2014), *w*Mel (Wu et al. 2004), *w*Ha (Ellegaard et al. 2013), and *w*No (Ellegaard et al. 2013).

Wolbachia gene extraction and phylogenetics

To determine phylogenetic relationships and estimate divergence times, we obtained the public *Wolbachia* Supergroup B genomes of: *w*AlbB that infects *Aedes albopictus* (Mavingui et al.

2012), *w*Pip_Pel that infects *Culex pipiens* (Klasson et al. 2008), *w*Pip_Mol that infects *Culex molestus* (Pinto et al. 2013), *w*No that infects *Drosophila simulans* (Ellegaard et al. 2013), and *w*VitB that infects *Nasonia vitripennis* (Kent et al. 2011), in addition to outgroup Supergroup A *w*Mel that infects *D. melanogaster* (Wu et al. 2004). These previously published genomes and the five *w*Mau-infected *D. mauritiana* genomes were annotated with Prokka v. 1.11, which identifies orthologs to known bacterial genes (Seemann 2014). To avoid pseudogenes and paralogs, we used only genes present in a single copy, and with no alignment gaps, in all of the genome sequences. Genes were identified as single copy if they uniquely matched a bacterial reference gene identified by Prokka v. 1.11. By requiring all orthologs to have identical length in all of the draft *Wolbachia* genomes, we removed all loci with indels. 143 genes, a total of 113,943 bp, met these criteria when comparing all of these genomes. However, when our analysis was restricted to the five *w*Mau genomes, our criteria were met by 694 genes, totaling 704,613 bp. Including *w*No with the five *w*Mau genomes reduced our set to 671 genes with 682,494 bp. We calculated the percent differences for the three codon positions within *w*Mau and between *w*Mau and *w*No.

We estimated a Bayesian phylogram of the *Wolbachia* sequences with RevBayes 1.0.8 under the GTR + Γ model, partitioning by codon position (Höhna et al. 2016). Four independent runs were performed, which all agreed.

We estimated a chronogram from the *Wolbachia* sequences using the absolute chronogram procedure implemented in Turelli et al. (2018). Briefly, we generated a relative relaxed clock chronogram with the GTR + Γ model with the root age fixed to 1 and the data partitioned by codon position. The relaxed clock branch rate prior was Γ(2,2). We used substitution-rate estimates of $\Gamma(7,7) \times 6.87\times10^{-9}$ substitutions/3rd position site/year to transform the relative chronogram into an absolute chronogram. This rate estimate was chosen so that the upper and lower credible intervals matched the posterior distribution estimated by Richardson et al. (2012), assuming 10 generations/year, normalized by their median estimate of 6.87×10^{-9} substitutions/3rd position site/year. (To assess the robustness of our conclusions to model assumptions, we also performed a strict-clock analysis and a relaxed-clock analysis with branch rate prior Γ(7,7).) For each analysis, four independent runs were performed, which all agreed. Our analyses all support *w*No as sister to *w*Mau.

We also estimated a relative chronogram for the host species using the procedure implemented in Turelli et al. (2018). Our host phylogeny was based on the same 20 nuclear genes used in Turelli et al. (2018): *aconitase, aldolase, bicoid, ebony, enolase, esc, g6pdh, glyp, glys, ninaE, pepck, pgi, pgm, pic, ptc, tpi, transaldolase, white, wingless* and *yellow*.

Analysis of *Wolbachia* and mitochondrial genomes

We looked for copy number variation (CNV) between *w*Mau and its closest relative, *w*No across the whole *w*No genome. Reads from the five infected *w*Mau lines were aligned to the *w*No reference (Ellegaard et al. 2013) with bwa 0.7.12 (Li and Durbin 2009). We calculated the normalized read depth for each alignment over sliding 1,000-bp windows by dividing the average depth in the window by the average depth over the entire *w*No genome. The results were plotted and visually inspected for putative copy number variants (CNVs). The locations of CNVs were specifically identified with ControlFREEC v. 11.5 (Boeva et al. 2012), using a ploidy of one and a window size of 1,000. We calculated *P*-values for each identified CNV with the Wilcoxon Rank Sum and the Kolmogorov-Smirnov tests implemented in ControlFREEC.

We used BLAST to search for pairs of WO phage loci in *w*Mau and *w*No genomes that are known to cause and rescue CI. These include the orthologous loci pairs WD0631-632 and *w*Pa_0282-283 that cause CI in *w*Mel and in *w*Pip, respectively (Beckmann and Fallon 2013; Beckmann et al. 2017; LePage et al. 2017). We also searched for the related *w*Pip *w*Pa_0294-295 pair (Beckmann et al. 2017). WD0631-632 and *w*Pa_0282-283 orthologs have been referred to as Type I *cifA-B* pairs, and *w*Pa_0294-295 as a Type IV *cifA-B* pair (LePage et al. 2017; Lindsey et al. 2018). Beckmann et al. (2017) refer to *w*Pa_0282-283 as *cidA-B* and *w*Pa_0294-295 as *cinA-B* based on their predicted function. *w*No_RS01055 and *w*No_RS01050 have been identified as a Type III *cifA-B* pair in the *w*No genome (LePage et al. 2017; Lindsey et al. 2018). No_RS01055 and *w*No_RS01050 are highly diverged from WD0631-632/*w*Pa_0282-283 orthologs and from *w*Pa_0294-295; however, this *w*No pair is more similar to *w*Pa_0294-295 in terms of protein domains, lacking a ubiquitin-like protease domain (Lindsey et al. 2018).

We found only the *w*No_RS01050/*w*No_RS01055 pair in *w*Mau genomes. The orthologs for *w*No_RS01050 and *w*No_RS01055 were extracted from our draft *w*Mau assemblies and aligned with MAFFT v. 7 (Katoh and Standley 2013). We compared these orthologs to

*w*No_RS01050/*w*No_RS01055 in *w*No and checked for single nucleotide variants (SNVs) among our *w*Mau assemblies.

D. mauritiana carry either the *ma*I mitochondrial haplotype, associated with *w*Mau infections, or the *ma*II haplotype (Ballard 2000a; James and Ballard 2000). To determine the mitochondrial haplotype of each *D. mauritiana* line, we assembled the mitochondrial genomes by down-sampling the reads by a factor of 100, then assembling with ABySS 2.0.2 using a *K* value of 71 for our data (150 bp reads) and 35 for the published data (76 bp reads) (Garrigan et al. 2014). Down-sampling reads prevents the nuclear genome from assembling but does not inhibit assembly of the mitochondrial genome, which has much higher coverage. We deemed the mitochondrial assembly complete if all protein-coding genes ($N = 13$) were present on the same contig and in the same order as in *D. melanogaster*. If the first attempt did not produce a complete mitochondrial assembly, we adjusted the down-sampling fraction until a complete assembly was produced for each line.

Annotated reference mitochondrial sequences for the *D. mauritiana* mitochondrial haplotypes *ma*I and *ma*II were obtained from Ballard et al. (2000b), and the thirteen protein-coding genes were extracted from our assemblies using BLAST and aligned to these references. The *ma*I and *ma*II reference sequences differ at 343 nucleotides over these protein-coding regions. We identified our lines as carrying the *ma*I haplotype if they differed by fewer than five nucleotides from the *ma*I reference and as *ma*II if they differed by fewer than five nucleotides from the *ma*II reference. None of our assemblies differed from both references at five or more nucleotides.

_w_Mau phenotypic analyses

Previous analyses have demonstrated that *w*Mau does not cause CI (Giordano et al. 1995). To check the generality of this result, we reciprocally crossed *w*Mau-infected *R31 D. mauritiana* with uninfected *R4* and measured egg hatch. Flies were reared under controlled conditions at 25°C for multiple generations leading up to the experiment. We paired 1–2-day-old virgin females with 1–2-day-old males in a vial containing spoons with cornmeal media and yeast paste. After 24 hr, pairs were transferred to new spoons, and this process was repeated for five days. Eggs on each spoon were given 24 hr at 25°C to hatch after flies were removed. To test for CI, we used nonparametric Wilcoxon tests to compare egg hatch between reciprocal crosses that

261 produced at least 10 eggs. All experiments were carried out at 25°C with a 12 light:12 dark
262 photoperiod.
263 To determine if *w*Mau generally enhances *D. mauritiana* fecundity, we assayed the
264 fecundity of *w*Mau-infected and uninfected *D. mauritiana* genotypes generated from reciprocal
265 crosses. Two *w*Mau-infected isofemale lines (*R31* and *R41*) were each reciprocally crossed to an
266 uninfected line (*R4*) to generate infected and uninfected F_1 females with similar genetic
267 backgrounds. The *w*Mau-infected and uninfected F_1 females were collected as virgins and placed
268 in holding vials. We paired 3–7-day-old females individually with an uninfected-*R4* male (to
269 stimulate oviposition) in vials containing a small spoon filled with standard cornmeal medium
270 and coated with a thin layer of yeast paste. We allowed females to lay eggs for 24 hours, after
271 which pairs were transferred to new vials. This was repeated for five days. At the end of each 24
272 hr period, spoons were frozen until counted. All experiments were carried out at 25°C with a 12
273 light:12 dark photoperiod.
274 We also assessed whether *D. mauritiana* genomic backgrounds modify *w*Mau fecundity
275 effects. We backcrossed *R31 w*Mau into the *R41 D. mauritiana* nuclear background by crossing
276 *R31* females with *R41* males. F_1 females were then backcrossed to *R41* males—this was repeated
277 for four generations to generate the $R41^{R31}$ genotypes (*w*Mau variant denoted by superscripts). A
278 similar approach was taken to generate $R31^{R41}$ genotypes. Each of these genotypes were then
279 reciprocally crossed to the uninfected *R4* line to generate infected and uninfected F_1 females with
280 controlled nuclear-genetic backgrounds. Upon emergence the *w*Mau-infected and uninfected F_1
281 females were collected as virgins and placed in holding vials. 3–7-day-old females were then
282 paired individually with an uninfected-*R4* male to stimulate oviposition, and egg-lay experiments
283 were carried out over five days as described above.
284 To determine if *w*Mau fecundity effects depend on host age, we measured egg lay of *w*Mau-
285 infected and uninfected *R31* genotypes over 24 days. Fast et al. (2011) carried out their
286 experiments over several weeks, which encouraged us to do the same. In an effort to more
287 closely mimic their approach, we generated tetracycline-treated *R31* genotypes (*R31-tet*) to
288 compare to infected *R31* genotypes. We fed flies 0.03% tetracycline concentrated medium for
289 four generations. We screened F_1 and F_2 individuals for *w*Mau using PCR, and we then fed flies
290 tetracycline food for two additional generations. In the fourth generation, we assessed *w*Mau titer
291 using qPCR to confirm that each genotype was cleared of *w*Mau infection. Following

tetracycline treatment, we placed *R31 D. mauritiana* males in food vials and allowed them to eat and defecate for 48 hours. We removed males, and *R31-tet* individuals were then placed in these same vials to reestablish the gut microbiome. Flies were given at least three more generations to avoid detrimental effects of tetracycline treatment on mitochondrial function (Ballard and Melvin 2007). To assay age effects on fecundity, we paired individual 6–7-day-old virgin *R31* (N = 30) and *R31-tet* (N = 30) females in bottles on yeasted apple-juice agar plates with an *R4* male to stimulate oviposition. Pairs were placed on new egg-lay plates every 24 hrs. After two weeks, we added one or two additional *R4* males to each bottle to replace any dead males and to ensure that females were not sperm limited as they aged.

To test for effects of *w*Mau on host fecundity, we used nonparametric Wilcoxon tests to compare the number of eggs laid between infected and uninfected females produced by reciprocally crossing *R31*, *R41*, $R3^{R41}$, and $R41^{R31}$ with uninfected *R4.* We also used Wilcoxon tests to assess variation in fecundity between infected-*R31* and uninfected-*R31-tet* genotypes across all 24 days of egg lay in our final experiment. We then assessed variation in *R31* and *R31-tet* egg lay across several ages to identify fecundity effects that vary with female age. Finally, we estimated the fitness parameter F in the standard discrete-generation model of CI to determine if *w*Mau fecundity effects underlie low frequency spread (Hoffmann et al. 1990; Turelli 1994). Pairs that laid fewer than 10 eggs across each experiment were excluded from analyses, but our results are robust to this threshold.

RESULTS

Wolbachia infection status

Out of 32 *D. mauritiana* lines that we analyzed, 11 were infected with *w*Mau *Wolbachia* (infection frequency = 0.34; binomial confidence interval: 0.19, 0.53). This new frequency estimate is not statistically different from a previous estimate (Giordano et al. 1995: infection frequency, 0.46; binomial confidence interval, (0.34, 0.58); Fisher's Exact Test, P = 0.293). These relatively low infection frequencies are consistent with theoretical expectations given that *w*Mau does not cause CI (Giordano et al. 1995; our data reported below), as *Wolbachia* that cause significant CI produce high equilibrium infection frequencies (Kriesner et al. 2016). Intermediate and apparently stable *w*Mau frequencies on Mauritius suggest that *w*Mau persists at a balance between positive effects on host fitness and imperfect maternal transmission.

Quantitative predictions, based on the idealized model of Hoffmann and Turelli (1997), are discussed below. The maintenance of *w*Mau is analogous to the persistence of other non-CI-causing *Wolbachia*, specifically *w*Suz in *D. suzukii* and *w*Spc in *D. subpulchrella* (Hamm et al. 2014; Conner et al. 2017; Turelli et al. 2018) and *w*Au in some Australian populations of *D. simulans* (Hoffmann et al. 1996; Kriesner et al. 2013).

Draft *w*Mau genome assemblies and comparison to *w*No

The five *w*Mau draft genomes we assembled were of very similar quality (Supplemental Table 2). N50 values ranged from 60,027 to 63,676 base pairs, and our assemblies varied in total length from 1,266,004 bases to 1,303,156 bases (Supplemental Table 2). Our BUSCO search found exactly the same genes in each draft assembly, and the presence/absence of genes in our *w*Mau assemblies was comparable to those in the complete genomes used as controls (Supplemental Table 3). In comparing our five *w*Mau draft genomes over 694 single-copy, equal-length loci comprising 704,613 bp, we found only one SNP. Four sequences (*R9*, *R31, R41* and *R60*) are identical at all 704,613 bp. *R29* differs from them at a single nucleotide, a nonsynonymous substitution in a locus which Prokka v. 1.11 annotates as "bifunctional DNA-directed RNA polymerase subunit beta/beta."

Comparing these five *w*Mau sequences to the *w*No reference (Ellegaard et al. 2013) over 671 genes with 682,494 bp, they differ by 0.068% overall, with equivalent divergence at all three codon positions (0.067%, 0.061%, and 0.076%, respectively).

***Wolbachia* phylogenetics**

As expected from the sequence comparisons, our Supergroup B phylogram places *w*Mau sister to *w*No (Figure 1A). This is consistent with previous analyses using fewer loci that placed *w*Mau (or *w*Ma in *D. simulans*) sister to *w*No (James and Ballard 2000; Zabalou et al. 2008; Toomey et al. 2013). Our chronogram (Figure 1B) estimates the 95% credible interval for the split between thee Supergroup B versus Subgroup A *Wolbachia* strains as 6 to 36 mya (point estimate = 16 mya). Reducing the variance on the substitution-rate-variation prior by using $\Gamma(7,7)$ rather than $\Gamma(2,2)$, changes the credible interval for the A-B split to 8 to 46 mya (point estimate = 21 mya); in contrast, a strict clock analysis produces a credible interval of 12 to 64 mya (point estimate = 31 mya). These estimates are roughly comparable to an earlier estimate that used a general

354 estimate for the synonymous substitution rate in bacteria (Ochman and Wilson 1987) and only
355 the *ftsZ* locus (59–67 mya, Werren et al. 1995). However, they are much lower than a recent
356 estimate based on comparative genomics (217 mya, Gerth and Bleidorn 2016). We discuss this
357 discrepancy below.
358

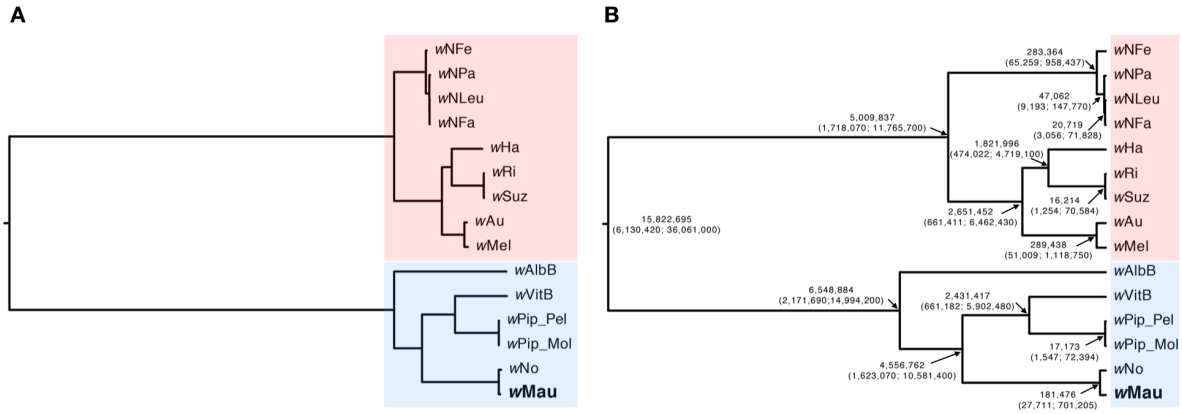

Figure 1. A) An estimated phylogram for various Supergroup A (red) and Supergroup B (blue) *Wolbachia* strains. All nodes have Bayesian posterior probabilities of 1. The phylogram shows significant variation in the substitution rates across branches, with long braches separating the A and B clades. **B)** An estimated chronogram for the same strains, with estimated divergence times and their confidence intervals at each node. To obtain these estimates, we generated a relative relaxed clock chronogram with the GTR + Γ model with the root age fixed to 1, the data partitioned by codon position, and with a Γ(2,2) branch rate prior. We used substitution-rate estimates of $\Gamma(7,7) \times 6.87 \times 10^{-9}$ substitutions/3rd position site/year to transform the relative chronogram into an absolute chronogram.

359
360 The observed divergence between *w*No and *w*Mau is consistent across all three codon
361 positions, similar to other recent *Wolbachia* splits (Turelli et al. 2018). Conversely, observed
362 divergence at each codon position generally varies across the chronogram, leading to inflation of
363 the *w*No-*w*Mau divergence point estimate of approximately 181,000 years (credible interval =
364 28,000 to 701,000 years; Figure 1B). To obtain a more accurate estimate, we calculated *w*No-
365 *w*Mau divergence using the observed third position pairwise divergence (0.077%, or 0.039%
366 from tip to MRCA) and Richardson et al. (2012)'s rate estimate. This approach produces a point
367 estimate of 57,000 years, with a credible interval of 30,000 to 135,000 years for the *w*No-*w*Mau
368 split, which seems plausible.
369
370

371 **Analysis of *Wolbachia* and mitochondrial genomes**

372 We looked for CNVs in *w*Mau relative to sister *w*No by plotting normalized read depth along the
373 *w*No genome. There were no differences in copy number among the *w*Mau variants, but
374 compared to *w*No, ControlFREEC identified four regions deleted from all *w*Mau that were
375 significant according to the Wilcoxon Rank Sum and Kolmogorov-Smirnov tests (Figure 2 and
376 Table 1). These deleted regions of the *w*Mau genomes include many genes, including many
377 phage-related loci, listed in Supplementary Table 4.

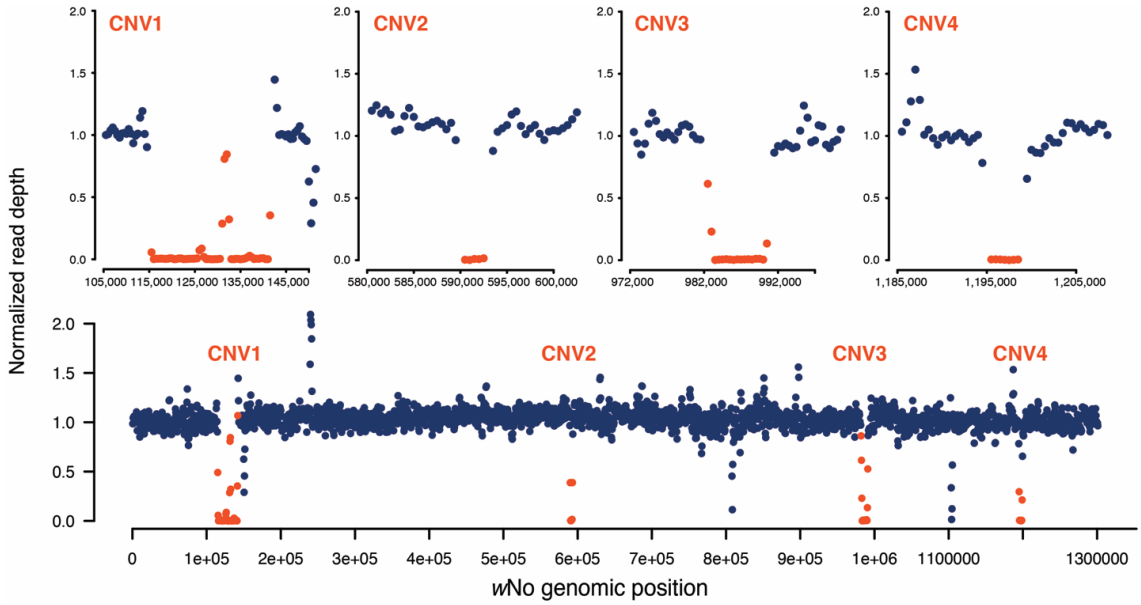

Figure 2. All *w*Mau variants share four large deletions, relative to sister *w*No. Top panel) The normalized read depth for *w*Mau *R60* plotted across the four focal regions of the *w*No reference genome; 10 kb of sequence surrounding regions are plotted on either side of each region. Bottom panel) The normalized read depth of *w*Mau *R60* plotted across the whole *w*No reference genome. Regions that do not contain statistically significant CNVs are plotted in dark blue, and regions with significant CNVs are plotted in red. All *w*Mau variants share the same CNVs, relative to *w*No.

378

379 To test the hypothesis that loci known to underlie CI are disrupted, we searched for pairs of
380 loci known to be associated with CI and found orthologs to the Type III gene pair
381 *w*No_RS01050-1055 in each of our draft assemblies, but we did not find orthologs to the
382 WD0631-632/*w*Pip_0282-283 pairs or to the *w*Pip_0294-295 pair. There were no variable sites
383 in the *w*No_RS01050-1055 gene regions among our five *w*Mau assemblies. All *w*Mau variants
384 share a one base pair deletion at base 1133 out of 2091 (amino acid 378) in the *w*No_RS01050

gene, relative to this same region in the *w*No genome. This frameshift introduces over ten stop codons, with the first at amino acid 388, likely making this predicted CI-causing protein nonfunctional. We also identified a nonsynonymous substitution in amino acid 264 of *w*No_RS01050 (*w*No codon ACA, Thr; *w*Mau codon AUA, Ile) and two SNVs in *w*No_RS01055: a synonymous substitution in amino acid 365 (*w*No codon GUC, *w*Mau codon GUU) and a nonsynonymous substitution in amino acid 397 (*w*No codon GCU, amino acid Ala; *w*Mau codon GAU, amino acid Asp). While one might intuitively expect selection to favor CI, disruption of CI in *w*Mau is consistent with theoretical analyses showing that selection within a host species does not act directly on the level of CI (Prout 1994; Turelli 1994; Haygood and Turelli 2009). Instead, selection should act to increase the product of *Wolbachia* effects on host fitness and the efficiency of maternal transmission (Turelli 1994).

Table 1. Copy number variants in *w*Mau relative to sister *w*No. Genomic positions are based on the *w*No reference. There were no CNVs among *w*Mau variants.

Start	End	Change	Wilcoxon Rank Sum Test	Kolmogorov-Smirnov Test
115,000	142,000	1 → 0	< 0.0001	< 0.0001
590,000	593,000	1 → 0	0.0027	0.0050
982,000	991,000	1 → 0	< 0.0001	< 0.0001
1,195,000	1,199,000	1 → 0	0.0005	0.0007

Of the *D. mauritiana* lines tested ($N = 9$), one line (uninfected-*R44*) carries the *ma*II mitochondrial haplotype, while the other eight carry *ma*I. The *ma*I and *ma*II references differ by 343 SNVs across the proteome, and *R44* differs from the *ma*II reference by 4 SNVs in the proteome. Four of our *ma*I lines (*R23*, *R29*, *R32*, and *R39*) are identical to the *ma*I reference, while three (*R31*, *R41*, and *R60*) have one SNV and one (*R9*) has two SNVs relative to *ma*I reference. One SNV is shared between *R9* and *R60*, but the other three SNVs are unique. Our results agree with past analyses that found *w*Mau is perfectly associated with the *ma*I mitochondrial haplotype (Ballard 2000a; James and Ballard 2000). The presence of *ma*II among the uninfected is interesting. In contrast to *ma*I, which is associated with introgression with *D. simulans* (Ballard 2000a; James and Ballard 2000), *ma*II appears as an outgroup on the mtDNA phylogeny of the *D. simulans* clade and is not associated with *Wolbachia* (Ballard 2000b, Fig. 5; James and Ballard 2000). Whether or not *Wolbachia* cause CI, if they are maintained by

selection-imperfect-transmission balance, we expect all uninfected flies to eventually carry the mtDNA associated with infected mothers (Turelli et al. 1992). We present a mathematical analysis of the persistence of *ma*II below.

Analysis of *w*Mau phenotypes

*w*Mau is reported to not cause CI (Giordano et al. 1995), but a single *w*Mau-infected *D. mauritiana* genotype displayed a four-fold increase in egg lay relative to an uninfected counterpart (Fast et al. 2011). We sought to determine whether *w*Mau can cause CI and whether *w*Mau generally increases *D. mauritiana* fecundity. In agreement with past analyses, we found no difference between the egg hatch of uninfected females crossed to *w*Mau-infected males (0.34 ± 0.23 SD, $N = 25$) and the reciprocal cross (0.29 ± 0.28 SD, $N = 24$). Contrary to previous results, we find no evidence for *w*Mau effects on *D. mauritiana* fecundity (Table 2 and Figure 3), regardless of host genetic backgrounds. Across both experiments assessing *w*Mau fecundity effects in their natural backgrounds (*R31* and *R41*), we counted 27,221 eggs and found no difference in the number of eggs laid by infected (mean = 238.20, SD = 52.35, $N = 60$) versus uninfected (mean = 226.82, SD = 67.21, $N = 57$) females over the five days of egg lay (Wilcoxon test, $W = 1540.5$, $P = 0.357$). Across both experiments that assessed *w*Mau fecundity effects in novel host backgrounds ($R31^{R41}$ and $R41^{R31}$), we counted 30,358 eggs and found no difference in the number of eggs laid by infected (mean = 253.30, SD = 51.99, $N = 60$) versus uninfected (mean = 252.67, SD = 63.53, $N = 60$) females over five days (Wilcoxon test, $W = 1869.5$, $P = 0.719$). [The mean number of eggs laid over 5 days, standard deviation (SD), sample size (N), and P-values from Wilcoxon tests are presented in Table 2 for all pairs.]

We sought to determine if *w*Mau fecundity effects depend on host age with a separate experiment that assessed egg lay over 24 days on apple-agar plates, similar to Fast et al. (2011). Across all ages, we counted 9,459 eggs and found no difference in the number of eggs laid by infected (mean = 187.70, SD = 168.28, $N = 27$) versus uninfected (mean = 156.29, SD = 138.04, $N = 28$) females (Wilcoxon test, $W = 409$, $P = 0.608$). This pattern was consistent across different age classes, suggesting age has little influence on *w*Mau fecundity effects (Table 3 and Figure 4). (The mean number of eggs laid, SD, N, and P-values from Wilcoxon tests are presented in Table 3 for all age classes.)

Table 2. *w*Mau does not significantly affect *D. mauritiana* fecundity in comparisons of paired infected (I) and uninfected (U) strains sharing host nuclear backgrounds. *N* is the number of females that produced the means and SDs. *P* values are for two-tailed Wilcoxon tests. Cytoplasm sources are denoted with superscripts for introgressed strains.

Strain	Mean eggs laid/5 days	SD	*N*	*P* value
R31I	210.97	55.22	30	0.901
R31U	213.50	58.43	28	
R41I	265.43	31.47	30	0.355
R41U	239.69	73.43	29	
$R31^{R41}I$	257.83	58.15	30	0.762
$R31^{R41}U$	254.67	48.72	30	
$R41^{R31}I$	248.77	45.54	30	0.433
$R41^{R31}U$	250.67	76.35	30	

441

Table 3. *w*Mau-infected (I) and uninfected (U) *D. mauritiana* females do not vary in egg lay at any of the age classes assessed. *N* is the number of females that produced the means and SDs. *P* values are for two-tailed Wilcoxon tests.

Egg lay period	Infection status	Mean eggs laid/5 days	SD	*N*	*P* value
Days 1-5	I	63.41	44.62	27	0.255
	U	79.46	57.98	26	
Days 6-10	I	64.80	38.43	20	0.172
	U	83.35	46.75	20	
Days 11-15	I	66.17	40.72	12	0.571
	U	61.0	49.09	14	
Days 16-20	I	47.56	21.45	9	0.555
	U	50.83	48.76	6	
Days 21-24	I	21.50	14.73	4	0.384
	U	29.0	16.51	4	

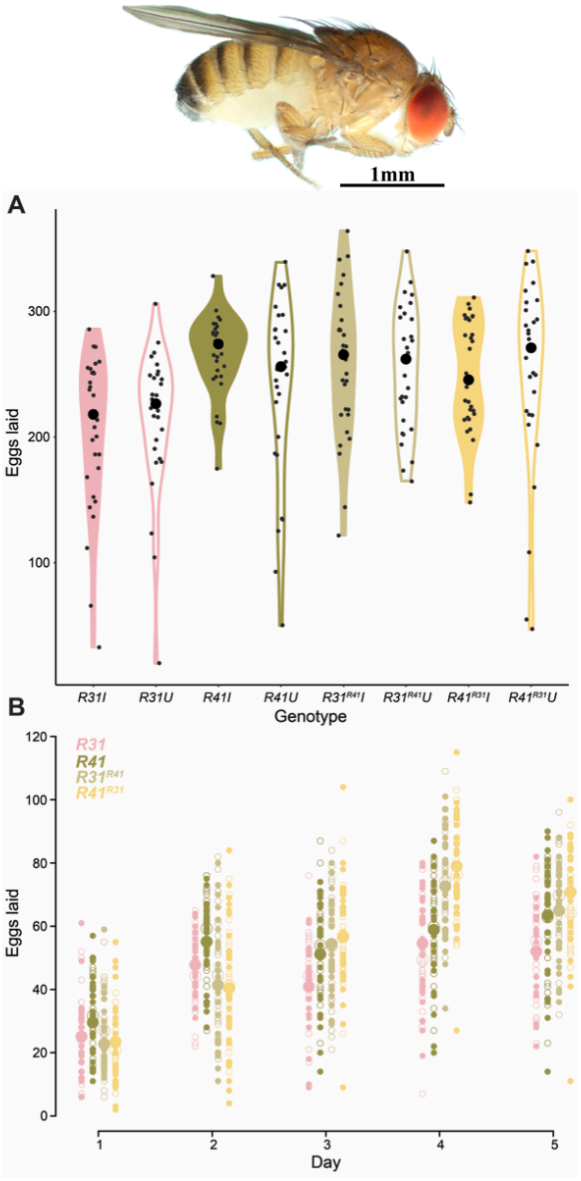

Figure 3. *w*Mau infections do not influence the number of eggs laid by *D. mauritiana* females (top) regardless of host genomic background. **A)** Violin plots of the number of eggs laid by *D. mauritiana* females over five days when infected with their natural *w*Mau variant (*R31I* and *R41I*), when infected with a novel *w*Mau variant ($R31^{R41}I$ and $R41^{R31}I$), and when uninfected (*R31U*, *R41U*, $R31^{R41}U$, and $R41^{R31}U$). Large black dots are medians, and small black dots are the eggs laid by each replicate over five days. **B)** The daily egg lay of these same infected (solid circles) and uninfected (open circles) strains is reported. Large circles are means of all replicates, and small circles are the raw data. Only days where females laid at least one egg are plotted. Cytoplasm sources are denoted by superscripts for the reciprocally introgressed strains.

442

With these data, we estimated the fitness parameter F in the standard discrete-generation model of CI (Hoffmann et al. 1990; Turelli 1994). Taking the ratio of replicate mean fecundity observed for *w*Mau-infected females to the replicate mean fecundity of uninfected females in naturally sampled *R31* and *R41 D. mauritiana* backgrounds, we estimated $F = 1.05$ (95% BC_a interval: 0.96, 1.16). Following reciprocal introgression of *w*Mau and host backgrounds (i.e., the $R31^{R41}$ and $R41^{R31}$ genotypes), we estimated $F = 1.0$ (95% BC_a interval: 0.93, 1.09). Finally, across all 24 days of our age-effects experiment, we estimated $F = 0.83$, 95% BC_a interval: 0.52, 1.34) for *R31*, which overlaps with our estimate of F for *R31* in our initial experiment (Table 4). Thus, and in support of our other analyses, there is little evidence that *w*Mau increases F to encourage spread. BC_a confidence intervals were calculated using the two-sample acceleration constant given by equation 15.36 of Efron and Tibshirani (1993). (Estimates of F and the associated BC_a confidence intervals are reported in Table 4 for different strains and age classes.) Our mathematical analysis below show that large fitness advantages associated with *w*Mau are inconsistent with the lack of CI and plausible levels of imperfect transmission.

Table 4. Estimates of the fitness parameter F in the standard discrete-generation model of CI indicate that *w*Mau fecundity effects have little influence on spread from low frequencies.

***w*Mau variant/age class**	***F***	**95% BC_a interval**
R31	0.988	(0.862, 1.137)
R41	1.107	(0.995, 1.265)
$R31^{R41}$	1.012	(0.911, 1.122)
$R41^{R31}$	0.992	(0.884, 1.143)
R31/Days 1-5	1.253	(0.845, 1.844)
R31/Days 6-10	1.286	(0.907, 1.836)
R31/Days 11-15	0.923	(0.531, 1.577)
R31/Days 16-20	1.069	(0.391, 2.11)
R31/Days 21-24	1.349	(0.60, 2.813)

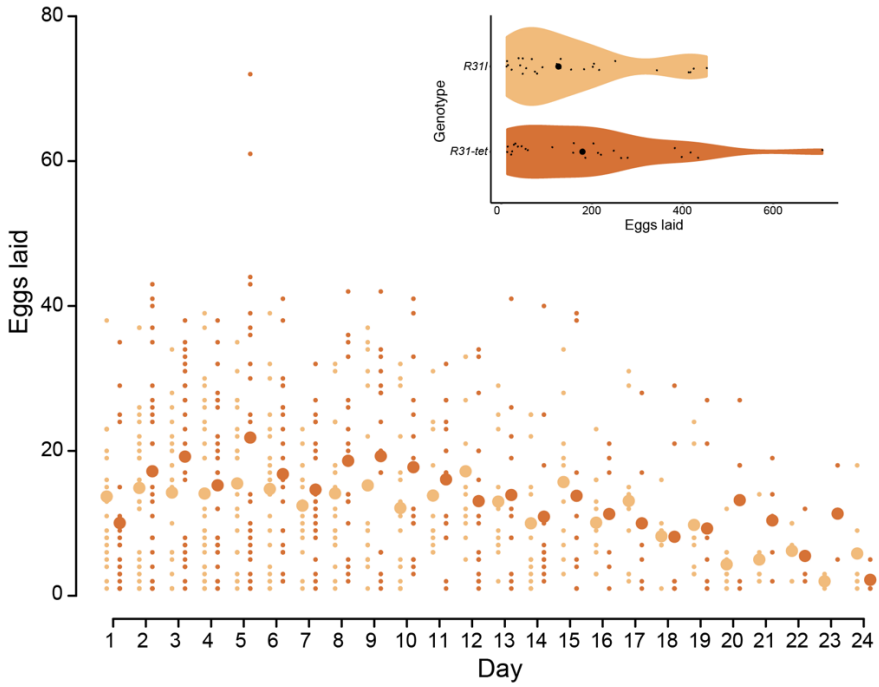

Figure 4. The mean number of eggs laid by infected *R31* (*R31I*, large tan dots) and *R31-tet* (large orange dots) are similar regardless of host age. Egg counts for each replicate are also plotted (small dots). Violin plots show egg lay across all ages for each genotype; large black circles are medians, and small black circles are the number of eggs laid by each replicate.

462
463

464 **Mathematical analyses of *Wolbachia* frequencies and mtDNA polymorphism**

465 If *Wolbachia* do not cause CI (or any other reproductive manipulation), their dynamics can be
466 approximated by a discrete-generation haploid selection model. Following Hoffmann and Turelli
467 (1997), we assume that the relative fecundity of *Wolbachia*-infected females is F, but a fraction
468 μ of their ova do not carry *Wolbachia*. Given our ignorance of the nature of *Wolbachia*'s
469 favorable effects, the F represents an approximation for all possible fitness benefits. If $F(1-\mu) >$
470 1, the equilibrium *Wolbachia* frequency among adults is

471

472 $$\hat{p} = 1 - \frac{\mu F}{F - 1}. \qquad (1)$$

473

474 Imperfect maternal transmisson has been documented for field-collected *D. simulans* infected
475 with *w*Ri (Hoffmann and Turelli 1988; Turelli and Hoffmann 1995; Carrington et al. 2011), *D.*

476 *melanogaster* infected with *w*Mel (Hoffmann et al. 1998) and *D. suzukii* infected with *w*Suz
477 (Hamm et al. 2014). The estimates range from about 0.01 to 0.1. In order for the equilibrium
478 *Wolbachia* frequency to be below 0.5, approximation (1) requires that the relative fecundity of
479 infected females satisfies
480
481 $$F < \frac{1}{1-2\mu}. \qquad (2)$$
482
483 Thus, even for μ as large a 0.15, which exceeds essentially all estimates of maternal transmission
484 failure from nature, *Wolbachia* can increase fitness by at most 43% and produce an equilibrium
485 frequency below 0.5. Conversely, (1) implies that a doubling of relative fecundity by *Wolbachia*
486 would produce an equilibrium frequency $1 - 2\mu$. If $\mu \leq 0.25$, consistent with all available data,
487 the predicted equilibrium implied by a fitness doubling significantly exceeds the observed
488 frequency of *w*Mau. Hence, a four-fold fecundity effect, as described by Fast et al. (2011), is
489 inconsistent with the frequency of *w*Mau in natural populations of *D. mauritiana*. Field estimates
490 of μ for *D. mauritiana* will provide better theory-based bounds on *w*Mau fitness effects.
491 As noted by Turelli et al. (1992), if *Wolbachia* is introduced into a population along with a
492 diagnostic mtDNA haplotype that has no effect on fitness, imperfect *Wolbachia* maternal
493 transmission implies that all infected and uninfected individuals will eventually carry the
494 *Wolbachia*-associated mtDNA, because all will have had *Wolbachia*-infected maternal ancestors.
495 We conjectured that a stable mtDNA polymorphism might be maintained if a *Wolbachia*-
496 associated mtDNA introduced by introgression is deleterious in its new nuclear background. We
497 refute our conjecture in Appendix 1. We show that the condition for *Wolbachia* to increase when
498 rare, $F(1 - \mu) > 1$, ensures that the native mtDNA will be completely displaced by the
499 *Wolbachia*-associated mtDNA, even if it lowers host fitness once separated from *Wolbachia*.
500 How fast is the mtDNA turnover, among *Wolbachia*-uninfected individuals, as a new
501 *Wolbachia* invades? This is easiest to analyze when the mtDNA introduced with *Wolbachia* has
502 no effect on fitness, so that the relative fitness of *Wolbachia*-infected versus uninfected
503 individuals is F, irrespective of the mtDNA haplotype of the uninfected individuals. As shown in
504 Appendix 1, the frequency of the ancestral mtDNA haplotype among uninfected individuals,
505 denoted r_t, declines as

$$r_{t+1} = r_t/[F(1-\mu)]. \quad (3)$$

Assuming $r_0 = 1$, recursion (3) implies that even if $F(1-\mu)$ is only 1.01, the frequency of the ancestral mtDNA haplotype should fall below 10^{-4} after 1000 generations. A much more rapid mtDNA turnover was seen as the CI-causing *w*Ri swept northward through California populations of *D. simulans* (Turelli et al. 1992; Turelli and Hoffmann 1995). Thus, it is anomolous that mtDNA haplotype *ma*II, which seems to be ancestral in *D. mauritiana* (Rousset and Solignac 1995; Ballard 2000a), persists among *Wolbachia*-uninfected *D. mauritiana*, given that all sampled *Wolbachia*-infected individuals carry *ma*I.

DISCUSSION

Understanding the genetic and phenotypic basis of *Wolbachia* frequencies in nature remains challenging and is particularly interesting for strains like *w*Mau that persist without inducing CI. Our results demonstrate intermediate and apparently stable *w*Mau frequencies on Mauritius, consistent with no CI. The lack of *w*Mau CI probably depends at least in part on the one-base-pair deletion that we identified in the *w*No_RS01050 gene. This mutation introduces a frameshift and more than ten stop codons, relative to sister *w*No, that we predict disrupt CI in *w*Mau. We also find no *w*Mau fecundity effects, indicating that *w*Mau must persist with unknown positive fitness effects balancing imperfect maternal transmission. We discuss the implications of our findings below.

*w*Mau is sister to *w*No and diverged from Supergroup A *Wolbachia* less than 100 mya

Our phylogenetic analyses place *w*Mau sister to *w*No, in agreement with past analyses using fewer data (James and Ballard 2000; Zabalou et al. 2008; Toomey et al. 2013). The relationships we infer agree with those from recently published phylograms (Gerth and Bleidorn 2016; Lindsey et al. 2018) (Figure 1A).

Depending on the prior we use for substitution-rate variation, we estimate that *w*Mau and other Supergroup B *Wolbachia* diverged from Supergroup A strains about 6-46 mya. This is roughly consistent with a prior estimate using only *ftsZ* (58 to 67 mya, Werren et al. 1995), but is inconsistent with a recent estimate using 179,763 bases across 252 loci (76–460 mya, Gerth and

Bleidorn 2016). There are several reasons why we question the Gerth and Bleidorn (2016) calibration. First, Gerth and Bleidorn (2016)'s chronogram placed *w*No sister to all other Supergroup B *Wolbachia*, in disagreement with their own phylogram (Gerth and Bleidorn 2016, Figure 3). In contrast, our phylogram and that of Lindsey et al. (2018) support *w*AlbB splitting from all other strains at this node. Second, the Gerth and Bleidorn (2016) calibration estimated the split between *w*Ri that infects *D. simulans* and *w*Suz that infects *D. suzukii* at 900,000 years. This estimate is nearly two orders of magnitude higher than the 11,000 year estimate of Turelli et al. (2018) who found 0.014% third position divergence between *w*Ri and *w*Suz (i.e., 0.007% along each branch) over 506,307 bases. Raychoudhury et al. (2009) and Richardson et al. (2012) both estimated a rate of about 7×10^{-9} substitutions/3rd position site/year between *Wolbachia* in *Nasonia* wasps and within *w*Mel, respectively. An estimate of 900,000 years requires a rate about 100 times slower, 7.8×10^{-11} substitutions/3rd position site/year, which seems implausible. Finally, using data kindly provided by Michael Gerth, additional analyses indicate that the third position rates required for the *Wolbachia* divergence times estimated by Gerth and Bleidorn (2016) between *Nomada flava* and *N. leucophthalma* (1.72×10^{-10}), *N. flava* and *N. panzeri* (3.78×10^{-10}) (their calibration point), and *N. flava* and *N. ferruginata* (4.14×10^{-10}) are each more than 10 times slower than those estimated by Raychoudhury et al. (2009) and Richardson et al. (2012), which seems unlikely. Our analyses suggest that the A-B group split occurred less than 100 mya.

Disruption of the CI phenotype generates relatively low infection frequencies

Across the 171 genes (139,902 bases) included in our phylogenetic analyses, the *w*Mau genomes were identical and differed from *w*No by only 0.072%. Across the coding regions we analyzed, we found few SNVs and no CNVs among *w*Mau variants. Our analyses did identify four large deletions shared by all *w*Mau genomes, relative to *w*No**.** Despite the close relationship between *w*Mau and *w*No, *w*No causes CI while *w*Mau does not (Giordano et al. 1995; Merçot et al. 1995; Rousset and Solignac 1995, our data). We searched for all pairs of loci known to cause CI and found only the *w*No_RS01050-1055 pair in both *w*Mau and *w*No genomes. All *w*Mau variants share a one-base-pair deletion in *w*No_RS01050, relative to this same region in *w*No. This mutation introduces a frameshift and more than ten stop codons that we predict disrupt CI in *w*Mau. Disruption of CI in *w*Mau is consistent with the prediction that selection does not directly

act on the intensity of CI (Prout 1994; Turelli 1994), and we predict that evidence of such disruption will be found in many other non-CI causing *Wolbachia* genomes—as has already been reported for a few other strains (Lindsey et al. 2018). Conversely, we predict that facultative *Wolbachia* lineages displaying loss of CI will be relatively recent. Importantly, non-CI *Wolbachia* have lower expected equilibrium infection frequencies than do CI-causing variants (Kriesner et al. 2016). This is consistent with a *w*Mau infection frequency of approximately 0.34 on Mauritius (Giordano et al. 1995; our data).

*w*Mau co-occurs with essentially the same mitochondrial haplotype as *w*Ma that infects *D. simulans* on Madagascar and elsewhere in Africa and the South Pacific (Rousset and Solignac 1995; Merçot and Poinsot 1998; Ballard 2000a; James and Ballard 2000; James et al. 2002; Ballard 2004), suggesting *w*Mau and *w*Ma may be the same strain infecting different host species following introgressive *Wolbachia* transfer (see below). The CI phenotypes of *w*Mau and *w*Ma are also more similar to one another than to *w*No, with only certain crosses between *w*Ma-infected *D. simulans* males and uninfected *D. simulans* females inducing CI (James and Ballard 2000). Polymorphism in the strength of CI induced by *w*Ma could result from host modification of *Wolbachia*-induced CI (Reynolds and Hoffmann 2002; Cooper et al. 2017), or from *Wolbachia* titer variation that influences the strength of CI and/or the strength of CI rescue by infected females. Alternatively, the single-base-pair deletion in the *w*No_RS01050 gene we observe in *w*Mau, or other mutations that influence CI strength, could be polymorphic in *w*Ma. *w*Ma infection frequencies are intermediate on Madagascar (infection frequency = 0.25, binomial confidence intervals: 0.14, 0.40; James and Ballard 2000), consistent with no CI, which suggests to us that crosses finding *w*Ma CI are likely anomalous or rare. Inclusion of *D. simulans* sampled from the island of Réunion in this estimate provides additional support for the conjecture that *w*Ma is non-CI-causing (infection frequency = 0.31, binomial confidence intervals: 0.20, 0.45; James and Ballard 2000). Unfortunately, *w*Ma-infected *D. simulans* were not available for us to genotype and phenotype.

Our genomic data indicate that *w*Mau may maintain an ability to rescue CI, as the *w*No_RS01055 gene is intact with only one nonsynonymous substitution relative to the same region in *w*No. *w*No_RS01055 is orthologous to WD0631 (*cifA*) in *w*Mel, recently shown to underlie CI rescue (Shropshire et al. 2018). *w*Ma seems to sometimes rescue CI, but conflicting patterns have been found, and additional experiments are needed to resolve this (Rousset and

599 Solignac 1995; Bourtzis et al. 1998; Merçot and Poinsot 1998; James and Ballard 2000; Merçot
600 and Poinsot 2003; Zabalou et al. 2008). Testing for CI rescue by *w*Mau- and *w*Ma-infected
601 females crossed with males infected with *w*No or other CI-causing strains, combined with
602 genomic analysis of CI loci in *w*Ma, will help resolve the relationship between these variants and
603 their CI phenotypes.
604
605 **_w_Mau does not influence *D. mauritiana* fecundity**
606 While selection does not directly act on the level of CI (Prout 1994; Turelli 1994; Haygood and
607 Turelli 2009), it does act to increase the product of *Wolbachia* effects on host fitness and the
608 efficiency of maternal transmission (Turelli 1994). Understanding the *Wolbachia* effects that
609 lead to spread from low frequencies and the persistence of non-CI causing *Wolbachia* is crucial
610 to explaining *Wolbachia* prevalence among insects and other arthropods. The four-fold fecundity
611 effect of *w*Mau reported by Fast et al. (2011) in *D. mauritiana* is inconsistent with our
612 experiments and with the intermediate infection frequencies observed in nature. We find no
613 *w*Mau effects on host fecundity, regardless of host background or female age. Our results are
614 consistent with an earlier analysis that assessed egg lay of a single genotype that found no *w*Mau
615 fecundity effects (Giordano et al. 1995), and with our mathematical analyses that indicate
616 *Wolbachia* can increase host fitness by at most 50% for reasonable estimates of μ.
617
618 **Introgressive *Wolbachia* transfer likely predominates in the *D. simulans* clade**
619 Hybridization and introgression in the *D. simulans* clade may have led to introgressive transfer
620 of *Wolbachia* among host species (Rousset and Solignac 1995), which has also been observed in
621 other *Drosophila* (Turelli et al. 2018) and in *Nasonia* wasps (Raychoudhury et al. 2009). While
622 *D. mauritiana* is singly infected by *w*Mau, *D. simulans* is infected by several strains that include
623 CI-causing *w*Ha and *w*No that often occur as double infections within individuals (O'Neill and
624 Karr 1990; Merçot et al. 1995; Rousset and Solignac 1995). *w*Ha and *w*No are similar to *w*Sh
625 and *w*Sn, respectively, that infect *D. sechellia* (Giordano et al. 1995; Rousset and Solignac
626 1995). *w*Ha and *w*Sh also occur as single infection in *D. simulans* and in *D. sechellia*,
627 respectively (Rousset and Solignac 1995). In contrast, *w*No almost always occurs alongside *w*Ha
628 in doubly infected *D. simulans* individuals (James et al. 2002), and *w*Sn seems to occur only in
629 individuals also infected by *w*Sh (Rousset and Solignac 1995). *D. simulans* has three distinct

630 mitochondrial haplotypes (*si*I, *si*II, *si*III) associated with *w*Au/*w*Ri (*si*II), *w*Ha/*w*No (*si*I), and
631 *w*Ma (*si*III). The *si*I haplotype is closely related to the *se* haplotype found with *w*Sh and *w*Sn in
632 *D. sechellia* (Ballard 2000b). *w*Ma co-occurs with the *si*III haplotype, which differs over its
633 protein-coding genes by a single base pair from the *ma*I mitochondrial haplotype carried by
634 *w*Mau-infected *D. mauritiana*—a second haplotype (*ma*II) is carried by only uninfected *D.*
635 *mauritiana* (Ballard 2000a; James and Ballard 2000).
636

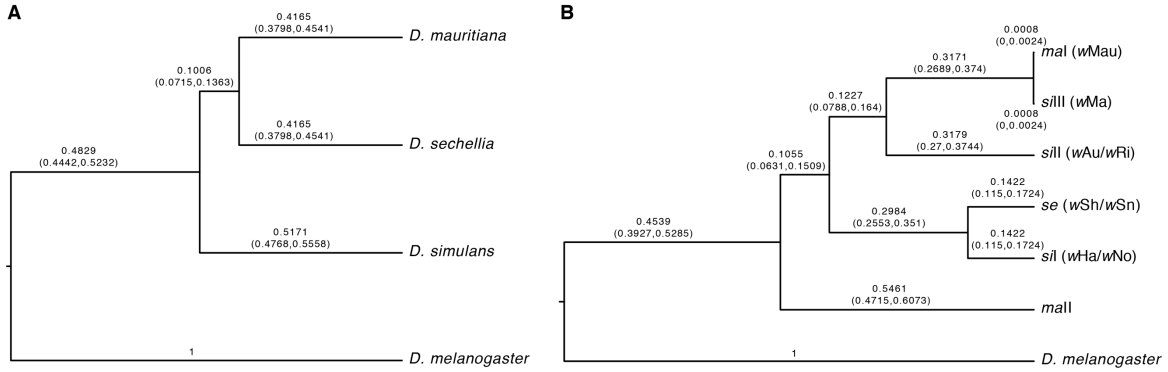

Figure 5. A) A nuclear relative chronogram. **B)** A mitochondrial relative chronogram with co-occurring *Wolbachia* strains listed in parentheses. See the text for an interpretation of the results, as noted the apparent resolution of the *D. simulans* clade is an artifact of the phylogenetic analysis.

637 The lack of whole *w*Ma genome data precludes us from confidently resolving the mode of
638 *w*Mau acquisition in *D. mauritiana*. However, mitochondrial relationships support the proposal
639 of Ballard (2000b) that *D. mauritiana* acquired *w*Mau and the *ma*I mitochondrial haplotype via
640 introgression from *w*Ma-infected *D. simulans* carrying *si*III. *D mauritiana* mitochondria are
641 paraphyletic relative to *D. sechellia* and *D. simulans* mitochondria (Solignac and Monnerot
642 1986; Satta and Takahata 1990; Ballard 2000a, 2000b), with *ma*I sister to *si*III and *ma*II outgroup
643 to all other *D. simulans*-clade haplotypes (see Figure 5). Of the nine genomes we assessed, all
644 but one (uninfected-*R44*) carry the *ma*I haplotype, and genotypes carrying *ma*I are both *w*Mau-
645 infected ($N = 5$) and uninfected ($N = 3$). While *w*Ma-infected *D. simulans* carry *si*III, *w*No-
646 infected *D. simulans* carry *si*I. We estimate that *w*Mau and *w*No diverged about 55,000 years
647 ago, with only 0.068% sequence divergence over 682,494 bp. Nevertheless, it seems implausible
648 that *w*No (versus *w*Ma) was transferred directly to *D. mauritiana* as this requires horizontal or
649 paternal transmission of *w*No into a *D. mauritiana* background already carrying the *ma*I

mitochondrial haplotype. Figure 5 shows relative chronograms for nuclear genes (A) and mtDNA haplotypes (B) for the *D. simulans* clade with *D. melanogaster* as the outgroup. Although the nuclear result suggests a confident phylogenetic resolution of the *D. simulans* clade, this is an artifact of the bifurcation structure imposed by the phylogenetic analysis. Population genetic analyses show a complex history of introgression and probable shared ancestral polymorphisms (Kliman et al. 2000) among these three species. Consistent with this, of the 20 nuclear loci we examined, 6 (*aconitase, aldolase, bicoid, ebony, enolase, ninaE*) supported *D. mauritiana* as the outgroup within the *D. simulans* clade, 7 (*glyp, pepck, pgm, pic, ptc, transaldolase, wingless*) supported *D. sechellia* as the outgroup, and 7 (*esc, g6pdh, glys, pgi, tpi, white, yellow*) supported *D. simulans*. With successive invasions of the islands and purely allopatric speciation, we expect the outgroup to be the island endemic that diverged first. Figure 5 indicates that the *ma*II haplotype diverged from the other mtDNA haplotypes roughly when the clade diverged, with the other haplotypes subject to a complex history of introgression and *Wolbachia*-associated sweeps, as described by Ballard (2000b).

Ballard (2000b) estimated that *si*III-*ma*I diverged about 4,500 years ago, which presumably approximates the date of the acquisition of *w*Mau (and *si*III, which became *ma*I) by *D. mauritiana*. This is surely many thousands of generations. As shown by our mathematical analyses (Eq. 3), the apparent persistence of the *ma*II mtDNA among *Wolbachia*-uninfected *D. mauritiana*—without its occurrence among infected individuals—is unexpected. More extensive sampling of natural *D. mauritiana* populations is needed to see if this unexpected pattern persists.

While paternal transmission has been observed in *D. simulans* (Hoffmann and Turelli 1988; Turelli and Hoffmann 1995), it seems to be very rare (Richardson et al. 2012; Turelli et al. 2018). *w*No almost always occurs in *D. simulans* individuals also infected with *w*Ha, complicating this scenario further. It is possible that horizontal or paternal transmission of *w*Ma or *w*No between *D. simulans* backgrounds carrying different mitochondrial haplotypes underlies the similarities of these strains within *D. simulans*, despite their co-occurrence with distinct mitochondria. Given the diversity of *Wolbachia* that infect *D. simulans*-clade hosts, and known patterns of hybridization and introgression among hosts (Garrigan et al. 2012; Brand et al. 2013; Garrigan et al. 2014; Matute and Ayroles 2014; Schrider et al. 2018), determining relationships among these *Wolbachia* and how *D. mauritiana* acquired *w*Mau will require detailed

phylogenomic analysis of nuclear, mitochondrial, and *Wolbachia* genomes in the *D. simulans* clade.

Conclusions

Disruption of CI contributes to intermediate *w*Mau infection frequencies on Mauritius that must be at balance between positive *w*Mau fitness effects and imperfect maternal transmission. The specific fitness effects underlying *Wolbachia* persistence remain elusive in this system as with most other non-CI *Wolbachia* infections. Analysis of fecundity effects and other phenotypes like nutritional supplementation and protection from harmful bacteria (Brownlie et al. 2009; Gupta et al. 2017; Kriesner and Hoffmann 2018), across a diversity of *Wolbachia*-infected hosts, will shed light of whether these phenotypes generally influence the spread and persistence of *Wolbachia*. Our mathematical analysis suggests that extreme fitness effects like those reported by Fast et al. (2011) are biologically implausible. Accurate estimates of *Wolbachia* effects on host fitness across host genomic backgrounds and abiotic environments, combined with estimates of *Wolbachia* maternal transmission in nature, are needed to explain the global *Wolbachia* pandemic and to improve the efficacy of transinfected *Wolbachia* as biocontrol agents.

AUTHOR CONTRIBUTIONS

MM performed the molecular and phenotypic work, participated in the design of the study, and contributed to the writing; WC performed the phylogenetic and genomic analyses and contributed to the writing; SR contributed to the molecular and phenotypic analyses and to the writing; JB performed the library preparation and contributed to the writing; MT contributed to the analyses, data interpretation, and writing; BSC designed and coordinated the study, contributed to the analyses and data interpretation, and drafted the manuscript. All authors gave final approval for publication.

ACKNOWLEDGMENTS

We thank Margarita Womack for sampling the *D. mauritiana* used in this study and Daniel Matute for sharing them. We thank Michael Gerth for sharing *Nomada* genomic data and Tim Wheeler for taking photos of *D. mauritiana*. Isaac Humble, Maria Kirby, and Tim Wheeler assisted with data collection. Michael Gerth, Michael Hague, and Amelia Lindsey provided

comments that improved earlier drafts of this manuscript. Computational resources were provided by the University of Montana Genomics Core. Research reported in this publication was supported by the National Institute Of General Medical Sciences of the National Institutes of Health (NIH) under Award Number R35GM124701 to B.S.C. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

DATA ACCESSIBILITY STATEMENT

All phenotypic data will be archived in DRYAD and all genetic data will be archived in GenBank at the time of acceptance.

REFERENCES

Baldo, L., N. A. Ayoub, C. Y. Hayashi, J. A. Russell, J. K. Stahlhut, and J. H. Werren. 2008 Insight into the routes of *Wolbachia* invasion: high levels of horizontal transfer in the spider genus *Agelenopsis* revealed by *Wolbachia* strain and mitochondrial DNA diversity. Molecular Ecolology 17**:**557–569.

Ballard, J. W. O. 2000a. When one is not enough: introgression of mitochondrial DNA in *Drosophila*. Molecular Biology and Evolution 17**:**1126–1130.

Ballard, J. W. 2000b. Comparative genomics of mitochondrial DNA in *Drosophila simulans*. Journal of Molecular Evolution 51:64–75.

Ballard, J. W. O. 2004. Sequential evolution of a symbioint inferred from the host: *Wolbachia* and *Drosophila simulans*. Molecular Biology and Evolution 21:428–442.

Ballard, J. W., and R. G. Melvin. 2007. Tetracycline treatment influences mitochondrial metabolism and mtDNA density two generations after treatment in *Drosophila*. Insect Molecular Biology 16**:**799–802.

Barton, N. H., and M. Turelli. 2011. Spatial waves of advance with bistable dynamics: cytoplasmic and genetic analogues of allee effects. American Naturalist 178**:**E48–E75.

Beckmann, J. F., and A. M. Fallon. 2013. Detection of the *Wolbachia* protein WPIP0282 in mosquito spermathecae: implications for cytoplasmic incompatibility. Insect Biochemistry and Molecular Biology 43**:**867–878.

Beckmann, J. F., J. A. Ronau and M. Hochstrasser. 2017. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. Nature Microbiology 2**:**17007.

Boeva, V., T. Popova, K. Bleakley, P. Chiche, J. Cappo, G. Schleiermacher, I. Janoueix-Lerosey, O. Delattre, E. Barillot. 2012. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 28**:**423–425.

Bourtzis, K., S. L. Dobson, H. R. Braig and S. L. O'Neill. 1998. Rescuing *Wolbachia* have been overlooked. Nature 391**:**852–853.

Brand, C. L., S. B. Kingan, L. Wu and D. Garrigan. 2013. A Selective Sweep across Species Boundaries in *Drosophila*. Molecular Biology and Evolution 30**:**2177–2186.

Brownlie, J. C., B. N. Cass, M. Riegler, J. J. Witsenburg, I. Iturbe-Ormaetxe, E. A. McGraw, S. L. O'Neill. 2009. Evidence for Metabolic Provisioning by a Common Invertebrate Endosymbiont, *Wolbachia pipientis*, during Periods of Nutritional Stress. PLoS Pathogens 5**:**e1000368.

Carrington, L. B., J. R. Lipkowitz, A. A. Hoffmann, and M. Turelli. 2011. A re-examination of *Wolbachia*-induced cytoplasmic incompatibility in California *Drosophila simulans*. PLoS ONE 6:e22565

Caspari, E., and G. S. Watson. 1959. On the evolutionary importance of cytoplasmic sterility in mosquitoes. Evolution 13:568–570.

Conner, W. R., M. L. Blaxter, G. Anfora, L. Ometto, O. Rota-Stabelli, and M. Turelli. 2017. Genome comparisons indicate recent transfer of *w*Ri-like *Wolbachia* between sister species *Drosophila suzukii* and *D. subpulchrella*. Ecology and Evolution 7:9391–9404.

Cooper, B. S., P. S. Ginsberg, M. Turelli and D. R. Matute. 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.

David, J., S. McEvey, M. Solignac and L. Tsacas. 1989. *Drosophila* communities on Mauritius and the ecological niche of *D. mauritiana* (Diptera, Drosophilidae). Revue De Zoologie Africaine – Journal of African Zoology 103:107–116.

789 Dean, M. D., and J. W. Ballard. 2004. Linking phylogenetics with population genetics to
790 reconstruct the geographic origin of a species. Molecular Phylogenetics and Evolution
791 32:998–1009.
792 Efron, B., and R. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New
793 York, NY.
794 Ellegaard, K. M., L. Klasson, K. Näslund, K. Bourtzis and S. G. E. Andersson. 2013.
795 Comparative Genomics of *Wolbachia* and the Bacterial Species Concept. PLoS Genetics
796 9:e1003381.
797 Fast, E. M., M. E. Toomey, K. Panaram, D. Desjardins, E. D. Kolaczyk, H. M. Frydman. 2011.
798 *Wolbachia* enhance *Drosophila* stem cell proliferation and target the germline stem cell
799 niche. Science 334:990–992.
800 Fisher, R. A. 1937. The wave of advance of advantageous genes. Annals of Eugenics 7:355–369.
801 Garrigan, D., S. B. Kingan, A. J. Geneva, P. Andolfatto, A. G. Clark, K. R. Thornton, D. C.
802 Presgraves. 2012. Genome sequencing reveals complex speciation in the *Drosophila*
803 *simulans* clade. Genome Research 22:1499–1511.
804 Garrigan, D., S. B. Kingan, A. J. Geneva, J. P. Vedanayagam and D. C. Presgraves. 2014.
805 Genome diversity and divergence in *Drosophila mauritiana*: multiple signatures of faster
806 X evolution. Genome Biology and Evolution 6:2444–2458.
807 Gerth, M., and C. Bleidorn. 2016. Comparative genomics provides a timeframe for *Wolbachia*
808 evolution and exposes a recent biotin synthesis operon transfer. Nature Microbiology
809 2:16241.
810 Giordano, R., S. L. O'Neill and H. M. Robertson. 1995. *Wolbachia* infections and the expression
811 of cytoplasmic incompatibility in *Drosophila sechellia* and *D. mauritiana*. Genetics
812 140:1307–1317.
813 Gloor, G. B., C. R. Preston, D. M. Johnsonschlitz, N. A. Nassif, R. W. Phillis, W. K. Benz, H.
814 M. Robertson, W. R. Engels. 1993. Type-1 repressors of P-element mobility. Genetics
815 135:81–95.
816 Gupta, V., R. B. Vasanthakrishnan, J. Siva-Jothy, K. M. Monteith, S. P. Brown, and P. F. Vale.
817 2017. The route of infection determines *Wolbachia* antibacterial protection in
818 *Drosophila*. Proceedings of the Royal Society B: Biological Sciences 284:20170809.
819 Hamm, C. A., D. J. Begun, A. Vo, C. C. R. Smith, P. Saelao, A. O. Shaver, J. Jaenike, M.
820 Turelli. 2014. *Wolbachia* do not live by reproductive manipulation alone: infection
821 polymorphism in *Drosophila suzukii* and *D. subpulchrella*. Molecular Ecology 23:4871–
822 4885.
823 Haygood, R., and M. Turelli. 2009. Evolution of incompatibility-inducing microbes in
824 subdivided host populations. Evolution 63:432–447
825 Hedges, L. M., J. C. Brownlie, S. L. O'Neill and K. N. Johnson. 2008. *Wolbachia* and virus
826 protection in insects. Science 322:702.
827 Hilgenboecker, K., P. Hammerstein, P. Schlattmann, A. Telschow and J. H. Werren. 2008. How
828 many species are infected with *Wolbachia*? A statistical analysis of current data. FEMS
829 Microbiology Letters 281:215–220.
830 Hoffmann, A. A., and M. Turelli. 1988. Unidirectional incompatibility in *Drosophila simulans*:
831 Inheritance, geographic variation and fitness effects. Genetics 119:435–444.
832 Hoffmann, A. A., and M. Turelli. 1997. Cytoplasmic incompatibility in insects. In *Influential*
833 *passengers: inherited microorganisms and arthropod reproduction*, edited by S. L.
834 O'Neill, J. H. Werren, and A. A. Hoffmann. Oxford University Press, pp. 42–80.

835 Hoffmann, A. A., D. Clancy and J. Duncan. 1996. Naturally-occurring *Wolbachia* infection in
836 *Drosophila simulans* that does not cause cytoplasmic incompatibility. Heredity 76**:**1–8.
837 Hoffmann, A. A., M. Hercus and H. Dagher. 1998. Population dynamics of the *Wolbachia*
838 infection causing cytoplasmic incompatibility in *Drosophila melanogaster*. Genetics
839 148:221–231.
840 Hoffmann, A. A., M. Turelli, and L. G. Harshman. 1990. Factors affecting the distribution of
841 cytoplasmic incompatibility in *Drosophila simulans*. Genetics 126**:**933–948.
842 Hoffmann, A. A., B. L. Montgomery, J. Popovici, I. Iturbe-Ormaetxe, P. H. Johnson, F. Muzzi,
843 M. Greenfield, M. Durkan, Y. S. Leong, Y. Dong, H. Cook, J. Axford, A. G. Callahan, N.
844 Kenny, C. Omodei, E. A. McGraw, P. A. Ryan, S. A. Ritchie, M. Turelli, and S. L.
845 O'Neill. 2011. Successful establishment of *Wolbachia* in *Aedes* populations to suppress
846 dengue transmission. Nature 476:454–457.
847 Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck,
848 F. Ronquist. 2016. RevBayes: Bayesian Phylogenetic Inference Using Graphical Models
849 and an Interactive Model-Specification Language. Systematic Biology. 65:726–736.
850 Huigens, M. E., R. P. de Almeida, P. A. Boons, R. F. Luck and R. Stouthamer. 2004. Natural
851 interspecific and intraspecific horizontal transfer of parthenogenesis-inducing *Wolbachia*
852 in *Trichogramma* wasps. Proceedings of the Royal Society B: Biological Sciences
853 271**:**509–515.
854 Jackman, S. D., B. P. Vandervalk, H. Mohamadi, J. Chu, S. Yeo, S. A. Hammond, G. Jahesh, H.
855 Khan, L. Coombe, R. L. Warren, I. Birol. 2017. ABySS 2.0: resource-efficient assembly
856 of large genomes using a Bloom filter. Genome Research 27**:**768–777.
857 James, A. C., and J. W. O. Ballard. 2000. Expression of cytoplasmic incompatibility in
858 *Drosophila simulans* and its impact on infection frequencies and distribution of
859 *Wolbachia pipientis*. Evolution 54**:**1661–1672.
860 James, A. C., M. D. Dean, M. E. McMahon and J. W. Ballard. 2002. Dynamics of double and
861 single *Wolbachia* infections in *Drosophila simulans* from New Caledonia. Heredity
862 88**:**182–189.
863 Jeyaprakash, A., and M. A. Hoy. 2000. Long PCR improves *Wolbachia* DNA amplification: wsp
864 sequences found in 76% of sixty-three arthropod species. Insect Molecular Biology
865 9**:**393–405.
866 Joshi, N. A., and J. N. Fass. 2011. Sickle: A sliding-window, adaptive, quality-based trimming
867 tool for FastQ files (Version 1.33). Available at https://github.com/najoshi/sickle.
868 Katoh, K., and D. M. Standley. 2013. MAFFT Multiple Sequence Alignment Software Version
869 7: Improvements in Performance and Usability. Molecular Biology and Evolution
870 30**:**772–780.
871 Kent, B. N., L. Salichos, J. G. Gibbons, A. Rokas, I. L. Newton, M. E. Clark, S. R. Bordenstein.
872 2011. Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*)
873 determined by targeted genome capture. Genome Biology and Evolution 3**:**209–218.
874 Klasson, L., T. Walker, M. Sebaihia, M. J. Sanders, M. A. Quail, A. Lord, S. Sanders, J. Earl, S.
875 L. O'Neill, N. Thomson, S. P. Sinkins, J. Parkhill. 2008. Genome evolution of *Wolbachia*
876 strain *w*Pip from the *Culex pipiens* group. Molecular Biology and Evolution 25**:**1877–
877 1887.
878 Klasson, L., J. Westberg, P. Sapountzis, K. Näslund, Y. Lutnaes, A. C. Darby, Z. Veneti, L.
879 Chen, H. R. Braig, and R. Garrett. 2009. The mosaic genome structure of the *Wolbachia*

*w*Ri strain infecting *Drosophila simulans*. Proceedings of the National Academy of Sciences 106**:**5725–5730.

Kliman, R. M., P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman, A. J. Berry, J. McCarter, J. Wakeley, and J. Hey. 2000. The population genetics of the origin and divergence of the *Drosophila simulans* species complex. Genetics 156:1913–1931.

Kriesner, P., and A. A. Hoffmann. 2018. Rapid spread of a *Wolbachia* infection that does not affect host reproduction in *Drosophila simulans* cage populations. Evolution 72:1475–1487.

Kriesner, P., W. R. Conner, A. R. Weeks, M. Turelli, and A. A. Hoffmann. 2016. Persistence of a *Wolbachia* infection frequency cline in *Drosophila melanogaster* and the possible role of reproductive dormancy. Evolution 70:979–997.

Kriesner, P., A. A. Hoffmann, S. F., Lee, M. Turelli, and A. R. Weeks. 2013. Rapid sequential spread of two *Wolbachia* variants in *Drosophila simulans*. PLoS Pathogens 9:e1003607.

Lachaise, D., J. R. David, F. Lemeunier, L. Tsacas and M. Ashburner. 1986. The reproductive relationships of *Drosophila sechellia* with *D. mauritiana*, *D. simulans*, and *D. melanogaster* from the Afrotropical region. Evolution 40**:**262–271.

Legrand, D., T. Chenel, C. Campagne, D. Lachaise and M. L. Cariou. 2011. Inter-island divergence within *Drosophila mauritiana*, a species of the *D. simulans* complex: Past history and/or speciation in progress? Molecular Ecology 20**:**2787–2804.

LePage, D. P., J. A. Metcalf, S. R. Bordenstein, J. On, J. I. Perlmutter, J. D. Shropshire, E. M. Layton, L. J. Funkhouser-Jones, J. F. Beckmann, S. R. Bordenstein. 2017. Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. Nature 543**:**243–247.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25**:**1754–1760.

Lindsey, A. R. I., D. W. Rice, S. R. Bordenstein, A. W. Brooks, S. R. Bordenstein, A. W. Brooks, S. R. Bordenstein, I. L. Newton. 2018. Evolutionary Genetics of Cytoplasmic Incompatibility Genes *cifA* and *cifB* in Prophage WO of *Wolbachia*. Genome Biology and Evolution 10**:**434–451.

Lindsey, A. R. I., J. H. Werren, S. Richards and R. Stouthamer. 2016. Comparative Genomics of a Parthenogenesis-Inducing *Wolbachia* Symbiont. Genes|Genomes|Genetics. 6:2113–2123.

Martinez, J., B. Longdon, S. Bauer, Y.-S. Chan, W. J. Miller, K. Bourtzia, L. Teixeira, F. M. Jiggins. 2014. Symbionts commonly provide broad spectrum resistance to viruses in insects: A comparative analysis of *Wolbachia* strains. PLoS Pathogens 10:e1004369.

Matute, D. R., and J. F. Ayroles. 2014. Hybridization occurs between *Drosophila simulans* and *D. sechellia* in the Seychelles archipelago. Journal of Evolutionary Biology 27**:**1057–1068.

Mavingui, P., C. Valiente Moro, V. Tran-Van, F. Wisniewski-Dye, V. Raquin, G. Minard, F. H. Tran, D. Voronin, Z. Rouy, P. Bustos, L. Lozano, V. Barbe, V. Gonzalez. 2012. Whole-genome sequence of *Wolbachia* strain *w*AlbB, an endosymbiont of tiger mosquito vector *Aedes albopictus*. Journal of Bacteriology 194**:**1840.

McDermott, S. R., and R. M. Kliman. 2008. Estimation of isolation times of the island species in the *Drosophila simulans* complex from multilocus DNA sequence data. PLoS One 3**:**e2442.

Merçot, H., and D. Poinsot. 1998. …and discovered on Mount Kilimanjaro. Nature 391**:**853.

926 Merçot, H., and D. Poinsot. 2003. *Wolbachia* transmission in a naturally bi-infected *Drosophila*
927 *simulans* strain from New-Caledonia. Entomologia Experimentalis et Applicata 86**:**97–
928 103.
929 Merçot, H., B. Llorente, M. Jacques, A. Atlan and C. Montchamp-Moreau. 1995. Variability
930 within the Seychelles cytoplasmic incompatibility system in *Drosophila simulans*.
931 Genetics 141**:**1015–1023.
932 Nunes, M. D., P. O. Wengel, M. Kreissl and C. Schlotterer. 2010. Multiple hybridization events
933 between *Drosophila simulans* and *Drosophila mauritiana* are supported by mtDNA
934 introgression. Molecular Ecology 19**:**4695–4707.
935 Ochman, H., and A. C. Wilson. 1987. Evolution in bacteria: evidence for a universal substitution
936 rate in cellular organisms. Journal of Molecular Evolution 26:74–86.
937 O'Neill, S. L., R. Giordano, A. M. Colbert, T. L. Karr and H. M. Robertson. 1992. 16S rRNA
938 phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic
939 incompatibility in insects. Proceedings of the National Academy of Sciences 89**:**2699–
940 2702.
941 O'Neill, S. L., and T. L. Karr. 1990. Bidirectional incompatibility between conspecific
942 populations of *Drosophila simulans*. Nature 348**:**178–180.
943 Pinto, S. B., K. Stainton, S. Harris, Z. Kambris, E. R. Sutton, M. B. Bonsall, J. Parkhill, S. P.
944 Sinkins. 2013. Transcriptional regulation of *Culex pipiens* mosquitoes by *Wolbachia*
945 influences cytoplasmic incompatibility. PLoS Pathogens 9**:**e1003647.
946 Poinsot, D., K. Bourtzis, G, Markakis, C. Savakis, and H. Merçot. 1998. Wolbachia transfer from
947 *Drosophila melanogaster* into *D. simulans*: host effect and cytoplasmic incompatibility
948 relationships. Genetics 150:227–237.
949 Prout, T. 1994. Some evolutionary possibilities for a microbe that causes incompatibility in its
950 host. Evolution 48**:**909–911.
951 Raychoudhury, R., L. Baldo, D. C. S. G. Oliveira, and J. H. Werren. 2009. Modes of acquisition
952 of *Wolbachia*: horizontal transfer, hybrid introgression, and codivergence in the *Nasonia*
953 species complex. Evolution 63:165–183.
954 Reynolds, K. T., and A. A. Hoffmann. 2002. Male age, host effects and the weak expression or
955 non-expression of cytoplasmic incompatibility in *Drosophila* strains infected by
956 maternally transmitted *Wolbachia*. Genetics Research 80**:**79–87.
957 Richardson, M. F., L. A. Weinert, J. J. Welch, R. S. Linheiro, M. M. Magwire, F. M. Jiggins, C.
958 M. Bergman. 2012. Population genomics of the *Wolbachia* endosymbiont in *Drosophila*
959 *melanogaster*. PLoS Genetics 8**:**e1003129.
960 Rousset, F., and M. Solignac. 1995. Evolution of single and double *Wolbachia* symbioses during
961 speciation in the *Drosophila simulans* complex. Proceedings of the National Academy of
962 Sciences 92**:**6389–6393.
963 Satta, Y. and N. Takahata. 1990. Evolution of *Drosophila* mitochondrial DNA and the history of
964 the *melanogaster* subgroup. Proceedings of the National Academy of Sciences of the
965 United States of America 87:9558–9562.
966 Schmidt, T. L., N. H. Barton, G. Rašić, A. P. Turley, B. L. Montgomery, I. Iturbe-Ormaetxe, P.
967 E. Cook, P. A. Ryan, S. A. Ritchie, A. A. Hoffmann, S. L. O'Neill, and M. Turelli. 2017.
968 Local introduction and heterogeneous spatial spread of dengue-suppressing *Wolbachia*
969 through an urban population of *Aedes aegypti*. PLoS Biology 15:e2001894.

970 Schrider, D. R., J. Ayroles, D. R. Matute and A. D. Kern. 2018. Supervised machine learning
971 reveals introgressed loci in the genomes of *Drosophila simulans* and *D. sechellia*. PLoS
972 Genetics 14**:**e1007341.
973 Schuler, H., K. Köppler, S. Daxböck-Horvath, B. Rasool, S. Krumböck, D. Schwarz, T. S.
974 Hoffmeister, B. C. Schlick-Steiner, F. M. Steiner, A. Telschow, C. Stauffler, W.
975 Arthofer, M. Riegler. 2016. The hitchhiker's guide to Europe: the infection dynamics of
976 an ongoing *Wolbachia* invasion and mitochondrial selective sweep in *Rhagoletis cerasi*.
977 Molecular Ecology 25**:**1595–1609.
978 Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30**:**2068-2069.
979 Shi, M., V. L. White, T. Schlub, J.-S. Eden, A. A. Hoffmann, E. C. Holmes. 2018. No detectable
980 effect of *Wolbachia w*Mel on the prevalence and abundance of the RNA virome of
981 *Drosophila melanogaster*. Proceedings of the Royal Society B: Biological Sciences
982 285:20181165.
983 Shropshire, J. D., J. On, E. M. Layton, H. Zhou and S. R. Bordenstein. 2018. One prophage WO
984 gene rescues cytoplasmic incompatibility in *Drosophila melanogaster*. Proceedings of the
985 National Academy of Sciences. 115:4987–4991.
986 Simao, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva and E. M. Zdobnov. 2015.
987 BUSCO: assessing genome assembly and annotation completeness with single-copy
988 orthologs. Bioinformatics 31**:**3210–3212.
989 Solignac, M. and M. Monnerot. 1986. Race formation, speciation, and introgression within
990 *Drosophila simulans*, *D. mauritiana*, and *D. schellia* inferred from mitochondrial DNA
991 analysis. Evolution 40:531–539.
992 Sutton, E. R., S. R. Harris, J. Parkhill and S. P. Sinkins. 2014. Comparative genome analysis of
993 *Wolbachia* strain *w*Au. BMC Genomics 15**:**928.
994 Taylor, M. J., D. Voronin, K. L. Johnston and L. Ford. 2013. *Wolbachia* filarial interactions. Cell
995 Microbiology 15**:** 520–526.
996 Team, R. C. 2015. R: A language and environment for statistical computing, *R Foundation for*
997 *Statistical Computing*, Vienna, Austria.
998 Teixeira, L., A. Ferreira and M. Ashburner. 2008. The bacterial symbiont *Wolbachia* induces
999 resistance to RNA viral infections in *Drosophila melanogaster*. PLoS Biology
1000 6**:**e1000002.
1001 Toomey, M. E., K. Panaram, E. M. Fast, C. Beatty and H. M. Frydman. 2013. Evolutionarily
1002 conserved *Wolbachia*-encoded factors control pattern of stem-cell niche tropism in
1003 *Drosophila* ovaries and favor infection. Proceedings of the National Academy of
1004 Sciences 110**:**10788–10793.
1005 Turelli, M. 1994. Evolution of incompatibility-inducing microbes and their hosts. Evolution
1006 48:1500–1513.
1007 Turelli, M. and N. H. Barton. 2017. Deploying dengue-suppressing *Wolbachia*: robust models
1008 predict slow but effective spatial spread in *Aedes aegypti*. Theoretical Population Biology
1009 115:45–60.
1010 Turelli, M. and A. A. Hoffmann. 1995. Cytoplasmic incompatibility in *Drosophila simulans*:
1011 dynamics and parameter estimates from natural populations. Genetics 140:1319–1338.
1012 Turelli, M., A. A. Hoffmann, and S. W. McKechnie. 1992. Dynamics of cytoplasmic
1013 incompatibility and mtDNA variation in natural *Drosophila simulans* populations.
1014 Genetics 132:713–723.

Turelli, M., B. S. Cooper, K. M. Richardson, P. S. Ginsberg, B. Peckenpaugh, C. X. Antelope, K. J. Kim, M. R. May, A. Abrieux, D. A. Wilson, M. J. Bronski, B. R. Moore, J. J. Gao, M. B. Eisen, J. C. Chiu, W. R. Conner, A. A. Hoffmann. 2018. Rapid Global Spread of *w*Ri-like *Wolbachia* across Multiple *Drosophila*. Current Biology 28**:**963–971.

Weeks, A. R., M. Turelli, W. R. Harcombe, K. T. Reynolds, and A. A. Hoffmann. 2007. From parasite to mutualist: rapid evolution of *Wolbachia* in natural populations of *Drosophila*. PLoS Biology 5**:**e114.

Weinert, L. A., E. V. Araujo-Jnr, M. Z. Ahmed and J. J. Welch. 2015. The incidence of bacterial endosymbionts in terrestrial arthropods. Proceedings of the Royal Society of London B: Biological Sciences 282**:**20150249.

Werren, J. H., and D. M. Windsor. 2000. *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? Proceedings of the Royal Society B: Biological Sciences 267**:**1277–1285.

Werren, J. H., W. Zhang and L. R. Guo. 1995. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. Proceedings of the Royal Society B-Biological Sciences 261**:**55–63.

Wu, M., L. V. Sun, J. Vamathevan, M. Riegler, R. Deboy, J. C. Brownlie, E. A. McGraw, W. Martin, C. Esser, N. Ahmadinejad. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. PLoS Biology 2**:**e69.

Zabalou, S., A. Apostolaki, S. Pattas, Z. Veneti, C. Paraskevopoulos, I. Livadaras, G. Markakis, T. Brissac, H. Merçot, K. Bourtzis. 2008. Multiple rescue factors within a *Wolbachia* strain. Genetics 178**:**2145–2160.

Zug, R., and P. Hammerstein. 2012. Still a host of hosts for *Wolbachia*: Analysis of recent data suggests that 40% of terrestrial arthropod species are infected. PLoS ONE 7:e38544.

Appendix 1. Mathematical analyses of mtDNA and *Wolbachia* dynamics

Our analysis follows the framework developed in Turelli et al. (1992), but is simplified by the lack of CI. We suppose that introgression introduces a cytoplasm carrying *Wolbachia* and a novel mtDNA haplotype, denoted B. Before *Wolbachia* introduction, we assume the population is monomorphic for mtDNA haplotype A. With imperfect maternal *Wolbachia* transmission, uninfected individuals will be produced with mtDNA haplotype B. Without horizontal or paternal transmission (which are very rare, Turelli et al. 2018), all *Wolbachia*-infected individuals will carry mtDNA haplotype B. Once *Wolbachia* is introduced, uninfected individuals can have mtDNA haplotype A or B. We assume that these three cytoplasmic types ("cytotypes") differ only in fecundity, and denote their respective fecundities F_I, F_A and F_B. Denote the frequencies of the three cytotypes among adults in generation t by $p_{I,t}$, $p_{A,t}$ and $p_{B,t}$, with $p_{I,t} + p_{A,t} + p_{B,t} = 1$. Without CI, the frequency dynamics are

$$p_{I,t+1} = \frac{p_{I,t}F_I(1-\mu)}{\overline{F}},\ p_{A,t+1} = \frac{p_{A,t}F_A}{\overline{F}},\ \text{and}\ p_{B,t+1} = \frac{p_{B,t}F_B + p_{I,t}\mu F_I}{\overline{F}},\ \text{with} \qquad \text{(A1)}$$

$$\overline{F} = F_I p_{I,t} + F_A p_{A,t} + F_B p_{B,t}. \qquad \text{(A2)}$$

If the uninfected population is initially monomorphic for mtDNA haplotype A, the *Wolbachia* infection frequency will increase when rare if and only if

$$F_I(1-\mu) > F_A. \qquad \text{(A3)}$$

Turelli et al. (1992) showed that if a CI-causing *Wolbachia* is introduced with a cytoplasm that contains a novel mtDNA haplotype B, which has no effect on fitness, *Wolbachia*-uninfected individuals will eventually all carry haplotype B. This follows because eventually all uninfected individuals have *Wolbachia*-infected maternal ancestors. This remains true for non-CI-causing *Wolbachia* that satisfy (A3). However, we conjectured that if the introduced B mtDNA is deleterious in the new host nuclear background, i.e., $F_A > F_B$, a stable polymorphism might be maintained for the alternative mtDNA haplotypes. The motivation was that imperfect maternal transmission seemed analogous to migration introducing a deleterious allele into an "island" of uninfected individuals. To refute this conjecture, consider the equilibria of (A1) with

$$F_I > F_A \geq F_B. \quad \text{(A4)}$$

If all three cytotypes are to be stably maintained, we expect each to increase in frequency when rare. In particular, we expect the fitness-enhancing mtDNA haplotype A to increase when the population contains only infected individuals and uninfected individuals carrying the deleterious *Wolbachia*-associated mtDNA haplotype B. From (A1), $p_{A,t}$ increases when rare if and only if

$$F_A > \overline{F} = F_I p_{I,t} + F_B(1 - p_{I,t}). \quad \text{(A5)}$$

In the absence of haplotype A, we expect p_I to be at equilibrium between selection and imperfect maternal transmission, i.e.,

$$p_I = 1 - \frac{\mu F}{F - 1}, \quad \text{(A6)}$$

with $F = F_I/F_B$ (Hoffmann and Turelli 1997). Substituting (A6) into (A5) and simplifying, the condition for $p_{A,t}$ to increase when rare is

$$F_A(F_I - F_B) > F_I(1 - \mu)(F_I - F_B). \quad \text{(A7)}$$

By assumption (A4), $F_I > F_B$; hence (A7) contradicts condition (A3), required for initial *Wolbachia* invasion. Thus, simple selection on *Wolbachia*-uninfected mtDNA haplotypes cannot stably maintain an mtDNA polymorphism. The "ancestral" mtDNA haplotype A is expected to be replaced by the less-fit *Wolbachia*-associated haplotype B.

To understand the time scale over which this replacement occurs, let r_t denote the frequency of haplotype A among *Wolbachia*-uninfected individuals, i.e., $r_t = p_{A,t}/(p_{A,t} + p_{B,t})$. From (A1),

$$r_{t+1} = \frac{r_t F_A}{r_t F_A + (1 - r_t)F_B + \mu F_I[p_{I,t}/(1 - p_{I,t})]}. \quad \text{(A8)}$$

If we assume that the mtDNA haplotypes do not affect fitness, i.e., $F_A = F_B$, and that the *Wolbachia* infection frequency has reached the equilibrium described by (A6), (A8) reduces to

$$r_{t+1} = r_t/[F(1 - \mu)], \qquad \text{(A9)}$$

with $F = F_I/F_B$.

SUPPLEMENTAL INFORMATION

Supplemental Table 1. PCR and qPCR primers used in this study.

Reaction	Locus/target region	Forward primer	Reverse primer	Reference
PCR	*wsp: Wolbachia* surface protein	5'-TGGTCCAATAAGTGATGAAGAAAC-3'	5'-AAAAATTAAACGCTACTCCA-3'	Braig et. al. 1998
PCR	*2L Control*: rDNA	5'-TGCAGCTATGGTCGTTGACA-3'	5'-ACGAGACAATAATATGTGGTGCTG-3'	Designed here
qPCR	*wsp: Wolbachia* surface protein	5'-CATTGGTGTTGGTGTTGGTG-3'	5'-ACCGAAATAACGAGCTCCAG-3'	Newton and Sheehan 2014
qPCR	*Rpl32:* rDNA	5'-CCGCTTCAAGGGACAGTATC-3'	5'-CAATCTCCTTGCGCTTCTTG-3'	Newton and Sheehan 2014

Supplemental Table 2. *w*Mau assembly statistics.

Genotype	Scaffold count	N50 (bp)	Longest scaffold (bp)	Total length (bp)
R9	36	60,027	169,305	1,266,004
R29	36	61,106	169,295	1,277,467
R31	39	63,676	169,381	1,272,847
R41	42	61,106	170,537	1,303,156
R60	38	63,156	221,751	1,282,564

1157

Supplemental Table 3. Near-universal, single-copy proteobacteria genes (out of 221) found using BUSCO v. 3.0.0.

Genome	Complete	Duplicated	Fragment	Absent
*w*Ri	179	1	2	39
*w*Mel	179	1	2	39
*w*Au	180	1	2	38
*w*Ha	178	1	3	39
*w*No	180	1	4	36
R9	180	1	4	36
R29	180	1	4	36
R31	180	1	4	36
R41	180	1	4	36
R60	180	1	4	36

1158
1159

Supplementary Table 4: Genes present in regions deleted in *w*Mau relative to *w*No. Genes predicted to be pseudogenized in *w*No are shaded grey.

Accession number	Name
Deletion 1 (115,000-142,000):	
wNO_RS00550	Hypothetical protein
wNO_RS06015	Ankyrin repeat domain protein
wNO_RS00560	Pseudo IS256 family transposase, frameshifted
wNO_RS00565	Recombinase family protein
wNO_RS00570	DUF2924 domain-containing protein
wNO_RS00575	Ankyrin repeat domain-containing protein
wNO_RS00580	Ankyrin repeat domain-containing protein
wNO_RS00585	Phage tail protein
wNO_RS00590	Baseplate assembly protein GpJ
wNO_RS00595	Pseudo baseplate assembly protein W, frameshifted
wNO_RS00600	Hypothetical protein
wNO_RS00605	Phage baseplate assembly protein V
wNO_RS00610	Hypothetical protein
wNO_RS00615	Putative minor tail protein Z
wNO_RS00620	Hypothetical protein
wNO_RS00625	Minor capsid protein E
wNO_RS00630	Head decoration protein
wNO_RS00635	S49 family peptidase
wNO_RS00640	Pseudo phage portal protein, frameshifted
wNO_RS00645	Phage head stabilizing protein GpW
wNO_RS00650	Phage terminase large subunit family protein
wNO_RS00655	Ankyrin repeat domain-containing protein
wNO_RS00660	Hypothetical protein
wNO_RS00665	Hypothetical protein
wNO_RS00670	Sigma-70 family RNA polymerase sigma factor
wNO_RS00675	ATP-binding protein

wNO_RS00680	IS110 family transposase
Deletion 2 (590,000-593,000):	
wNO_RS02645	XRE family transcriptional regulator
wNO_RS02650	Hypothetical protein
wNO_RS02655	Hypothetical protein
wNO_RS02660	XRE family transcriptional regulator
Deletion 3 (982,000-991,000):	
wNO_RS04690	Group II intron reverse transcriptase/maturase
wNO_RS06355	Hypothetical protein
wNO_RS04695	Pseudo hypothetical protein, partial
wNO_RS04700	Pseudo cell filamentation protein Fic, partial
wNO_RS04705	Hypothetical protein
wNO_RS04710	DNA methylase
wNO_RS04715	Hypothetical protein
wNO_RS04720	Ankyrin repeat domain-containing protein
wNO_RS04725	Phage terminase large subunit family protein
wNO_RS04730	Phage head stabilizing protein GpW
wNO_RS04735	DUF1016 domain-containing protein
Deletion 4 (1,195,000-1,199,000):	
wNO_RS06095	Ankyrin repeat domain protein
wNO_RS05590	Rpn family recombination-promoting nuclease/putative transposase