Introgressive and horizontal acquisition of Wolbachia by *Drosophila yakuba*-clade hosts and horizontal transfer of incompatibility loci between distantly related Wolbachia

Brandon S. Cooper[*], Dan Vanderpool[§], William R. Conner[*], Daniel R. Matute[‡], Michael Turelli[†]

[*]Division of Biological Sciences
University of Montana
32 Campus Dr.
Missoula, MT 59812

[§]Department of Biology
Indiana University
1001 E. 3rd St.
Bloomington, IN 47405

[‡]Biology Department
University of North Carolina
250 Bell Tower Drive, Genome Sciences Building
Chapel Hill, NC 27510

[†]Department of Evolution and Ecology
University of California
One Shields Avenue
Davis, California 95616

Running title: Wolbachia and incompatibility loci transfer

Key words: cytoplasmic incompatibility, hybridization, introgression, horizontal gene transfer, host-microbe interactions, transposable elements, WO phage

Corresponding author: Brandon S. Cooper, (406) 243-5122, brandon.cooper@umontana.edu

26 February 2019

ABSTRACT

Maternally transmitted Wolbachia infect about half of insect species, yet the predominant mode(s) of Wolbachia acquisition remains uncertain. Species-specific associations could be old, with Wolbachia and hosts co-diversifying (i.e., cladogenic acquisition), or relatively young and acquired by horizontal transfer or introgression. The three *Drosophila yakuba*-clade hosts ((*D. santomea, D. yakuba*), *D. teissieri*) diverged about three million years ago and currently hybridize on Bioko and São Tomé, west African islands. Each species is polymorphic for nearly identical Wolbachia that cause weak cytoplasmic incompatibility (CI)–reduced egg hatch when uninfected females mate with infected males. *D. yakuba*-clade Wolbachia are closely related to *w*Mel, globally polymorphic in *D. melanogaster*. We use draft Wolbachia and mitochondrial genomes to demonstrate that *D. yakuba*-clade Wolbachia and mitochondrial phylogenies tend to follow host nuclear phylogenies; however, roughly half of *D. santomea* individuals, sampled both inside and outside of the São Tomé hybrid zone, have introgressed *D. yakuba* mitochondria. Both mitochondria and Wolbachia possess far more recent common ancestors than the bulk of the host nuclear genomes, precluding cladogenic Wolbachia acquisition. General concordance of Wolbachia and mitochondrial phylogenies suggests that horizontal transmission is rare, but varying relative rates of molecular divergence complicate time-based statistical tests. Loci that cause CI in *w*Mel are disrupted in *D. yakuba*-clade Wolbachia; however, a second set of loci predicted to cause CI are located in the same WO prophage region. These alternative CI loci seem to have been acquired horizontally from distantly related Wolbachia, with transfer mediated by flanking Wolbachia-specific ISWpi1 transposons.

INTRODUCTION

Endosymbiotic Wolbachia bacteria infect many arthropods (Bouchon *et al.* 1998; Hilgenboecker *et al.* 2008), including about half of all insect species (Werren and Windsor 2000; Weinert *et al.* 2015). Wolbachia often manipulate host reproduction, facilitating spread to high frequencies within host species (Laven 1951; Yen and Barr 1971; Turelli and Hoffmann 1991; Rousset *et al.* 1992; O'Neill *et al.* 1998; Weeks *et al.* 2007; Kriesner *et al*. 2016; Turelli *et al*. 2018). In Drosophila, reproductive manipulations include cytoplasmic incompatibility (CI) and male killing (Hoffmann *et al.* 1986; Hoffmann and Turelli 1997; Hurst and Jiggins 2000). CI reduces the egg hatch of uninfected females mated with Wolbachia-infected males, and recent work has demonstrated that WO prophage-associated loci cause CI (Beckmann and Fallon 2013; Beckmann *et al*. 2017; LePage *et al*. 2017; Beckmann *et al*. 2019). Although reproductive manipulations are common, some Wolbachia show little or no reproductive manipulation (*e.g*., *w*Mel in *D. melanogaster*, Hoffmann 1988; Hoffmann *et al.* 1994; Kriesner *et al*. 2016; *w*Mau in *D. mauritiana*, Giordano et al. 1995; Meany *et al.* 2018; *w*Au in *D. simulans*, Hoffmann *et al.* 1996; *w*Suz in *D. suzukii* and *w*Spc in *D. subpulchrella*, Hamm *et al*. 2014). These Wolbachia presumably spread by enhancing host fitness in various ways, with some support for viral protection, fecundity enhancement, and supplementation of host nutrition (Weeks *et al.* 2007; Teixeira *et al.* 2008; Hedges *et al.* 2008; Brownlie *et al.* 2009; Martinez *et al.* 2014; Gill *et al.* 2014; Moriyama *et al.* 2015). Better understanding of Wolbachia effects, transmission and evolution should facilitate using Wolbachia for biocontrol of human diseases by either transforming vector populations with virus-blocking Wolbachia (*e.g*., McMeniman *et al.* 2009; Hoffmann *et al.* 2011; Schmidt *et al*. 2017; Ritchie 2018) or using male-only releases of CI-causing Wolbachia to suppress vector populations (Laven 1967; O'Connor *et al*. 2012).

There is a burgeoning literature on Wolbachia frequencies and dynamics in natural populations (*e.g*., Kriesner *et al*. 2013; Kriesner *et al*. 2016; Cooper *et al*. 2017; Bakovic *et al*. 2018; Meany *et al.* 2018), but fewer studies elucidate the modes and time scales of Wolbachia acquisition by host species (O'Neill *et al*. 1992; Rousset and Solignac 1995; Huigens *et al*. 2004; Baldo *et al*. 2008; Raychoudhury *et al*. 2009; Ahmed *et al*. 2015; Schuler *et al*. 2016; Turelli *et al.* 2018). Sister hosts could acquire Wolbachia from their most recent ancestors. Such cladogenic acquisition seems to be the rule for the obligate Wolbachia found in filarial

nematodes (Bandi *et al*. 1998); and there are also examples in at least two insect clades, Nasonia wasps (Raychoudhury *et al.* 2009) and Nomada bees (Gerth and Bleidorn 2016). In contrast, infections can be relatively young and acquired through introgressive or horizontal transfer (Raychoudhury *et al.* 2009; Schuler *et al*. 2016; Conner *et al.* 2017; Turelli *et al.* 2018). Comparisons of host nuclear and mitochondrial genomes with the associated Wolbachia genomes enable discrimination among these three modes (Raychoudhury *et al.* 2009; Turelli *et al.* 2018). Concordant nuclear, mitochondrial, and Wolbachia cladograms—including consistent divergence-time estimates for all three genomes—support cladogenic acquisition and co-divergence of host and Wolbachia lineages. Concordant Wolbachia and mtDNA phylogenies and consistent Wolbachia and mitochondrial divergence-time estimates that are more recent than nuclear divergence support introgressive acquisition. In this case, mitochondrial and Wolbachia relationships may or may not recapitulate the host phylogeny. Finally, if Wolbachia diverged more recently than either nuclear or mitochondrial genomes, horizontal transfer (or paternal transmission; Hoffmann and Turelli 1988; Turelli and Hoffmann 1995) is indicated. This is often associated with discordance between host and Wolbachia phylogenies (O'Neill *et al.* 1992; Rousset and Solignac 1995; Turelli *et al.* 2018).

Introgressive Wolbachia acquisition may be common in Drosophila. About half of all closely related Drosophila species have overlapping geographical ranges and show pervasive evidence of reinforcement (Coyne and Orr 1989, 1997; Yukelevich 2012; Nosil 2013), indicating that hybridization must be common (Turelli *et al.* 2014). Several instances of sporadic contemporary hybridization and interspecific gene flow have been documented in the genus (Carson *et al.* 1989; Shoemaker *et al.* 1999; Jaenike *et al.* 2006; Kulathinal *et al.* 2009; Garrigan *et al.* 2012; Brand *et al.* 2013; Matute and Ayroles 2014; Lohse *et al.* 2015); but only two stable hybrid zones have been well described, and both involve *D. yakuba*-clade species. In west Africa, *D. yakuba* hybridizes with endemic *D. santomea* on the island of São Tomé, and with *D. teissieri* on the island of Bioko (Lachaise *et al.* 2000; Llopart *et al.* 2005; Comeault *et al.* 2016; Cooper *et al.* 2018). The ranges of *D. yakuba* and *D. teissieri* overlap throughout continental Africa, but contemporary hybridization has not been observed outside of Bioko (Cooper *et al.* 2018). Genomic analyses support both mitochondrial and nuclear introgression in the *D. yakuba* clade (Lachaise *et al.* 2000; Bachtrog *et al.* 2006; Llopart *et al.* 2014; Turissini and Matute 2017; Cooper *et al.* 2018), and the Wolbachia infecting all three species (*w*San, *w*Tei, and *w*Yak) are at

92 intermediate frequencies and identical with respect to commonly used typing loci (Lachaise *et al*.
93 2000; Charlat *et al*. 2004; Cooper *et al.* 2017).
94 Horizontal Wolbachia transmission has been repeatedly demonstrated since its initial
95 discovery by O'Neill *et al*. (1992) (*e.g*., Baldo *et al.* 2008; Schuler *et al.* 2016), and it may be
96 common in some systems (*e.g*., Huigens *et al.* 2000; Huigens *et al.* 2004; Ahmed *et al*. 2015; Li
97 *et al.* 2017). Phylogenomic analyses indicate recent horizontal Wolbachia transmission, on the
98 order of 5,000–27,000 years, among relatively distantly related Drosophila species of the *D.*
99 *melanogaster* species group (Turelli *et al.* 2018); but there is little evidence for non-maternal
100 transmission within Drosophila species (Richardson *et al.* 2012; Turelli *et al.* 2018). Horizontal
101 Wolbachia transfer could occur via a vector (Vavre *et al*. 2009; Ahmed *et al*. 2015) and/or via
102 shared food sources during development (Huigens *et al.* 2000; Li *et al*. 2017). Rare paternal
103 Wolbachia transmission has been documented in *D. simulans* (Hoffmann and Turelli 1988;
104 Turelli and Hoffmann 1995). However, the most common mode of Wolbachia acquisition by
105 Drosophila species (and most other hosts) remains unknown, and all three modes seem plausible
106 in the *D. yakuba* clade (Lachaise *et al*. 2000; Bachtrog *et al*. 2006; Cooper *et al*. 2017).
107 Distinguishing acquisition via introgression versus horizontal or paternal transmission
108 requires estimating phylogenies and relative divergence times of mtDNA and Wolbachia
109 (Raychoudhury *et al.* 2009; Conner *et al*. 2017; Turelli *et al*. 2018). To convert sequence
110 divergence to divergence time estimates requires understanding relative rates of divergence for
111 mtDNA, nuclear genomes and Wolbachia. If each genome followed a constant-rate molecular
112 clock, taxa with cladogenic Wolbachia transmission, such as Nasonia wasps (Raychoudhury *et*
113 *al.* 2009) and Nomada bees (Gerth and Bleidorn 2016), would provide reference rates for
114 calibration. However, like Langley and Fitch (1974), Turelli *et al*. (2018) found significantly
115 varying relative rates of Wolbachia and mtDNA divergence. This variation, which we document
116 more extensively, confounds attempts to understand Wolbachia acquisition, as discussed below.
117 The discovery of weak CI in the *D. yakuba clade* (Cooper *et al*. 2017) motivates detailed
118 comparative analysis of loci associated with CI (CI factors or *cifs*) in WO prophage regions of
119 Wolbachia genomes. Beckmann and Fallon (2013) first associated wPip_0282 and wPip_0283
120 proteins in *w*Pip-infected *Culex pipiens* with Wolbachia-modified sperm. Later work confirmed
121 that these proteins induce toxicity and rescue when expressed/co-expressed in *Saccharomyces*
122 *cerevisiae* (Beckmann *et al.* 2017), lending support to a toxin-antidote model of CI (Beckmann *et*

123 *al.* 2019; but see, Lindsey *et al.* 2018). Homologs of these genes in *w*Mel (WD0631 and
124 WD0632) recapitulate CI when transgenically expressed in *D. melanogaster* (LePage *et al.*
125 2017), and transgenic expression of WD0631 in *D. melanogaster* rescues CI (Shropshire *et al*.
126 2018). A distantly related WO-prophage-associated pair, present in *w*Pip, *w*Pip_0294 and
127 *w*Pip_0295, causes similar toxicity/rescue in *S. cerevisiae* (Beckmann *et al*. 2017), and causes CI
128 when placed transgenically into a *D. melanogaster* background (unpublished data presented by
129 Mark Hochstrasser at the 2018 Wolbachia meeting in Salem, MA). We adopt Beckmann *et al.*
130 (2019)'s nomenclature, which assigns names based on enzymatic activity of the predicted toxin
131 [deubiquitylase (DUB) and nuclease (Nuc)], with superscripts denoting focal Wolbachia strains
132 when needed. Specifically, we refer to *w*Pip_0282-*w*Pip_0283 and the wPip_0294-wPip_0295
133 pairs as *cidA*-*cidB*$^{w\mathrm{Pip}}$ and *cinA*-*cinB*$^{w\mathrm{Pip}}$, respectively; and we refer to WD0631-WD632 as *cidA*-
134 *cidB*$^{w\mathrm{Mel}}$. This distinguishes *cid* (CI-inducing deubiquitylating enzyme) from *cin* (CI-inducing
135 nuclease) pairs, with the predicted antidote and toxin denoted "A" and "B", respectively
136 (Beckmann *et al.* 2019). We acknowledge ongoing disagreement in the literature and direct
137 readers to Beckmann *et al.* (2019) and Lindsey *et al.* (2018) for more detailed discussions.
138 However, none of these debates on terminology or mechanism affect our findings.
139 Here, we use host and Wolbachia genomes from the *D. yakuba* clade to demonstrate
140 introgressive and horizontal Wolbachia acquisition. General concordance of mtDNA and
141 Wolbachia phylogenies indicates that horizontal acquisition is rare within this clade. However,
142 tests involving divergence-time estimates are complicated by varying relative rates and patterns
143 of Wolbachia, mtDNA, and nuclear sequence divergence, as illustrated by data from more
144 distantly related Drosophila (Clark *et al.* 2007). Finally, we demonstrate that *cid* loci underlying
145 CI in closely related *w*Mel (LePage *et al.* 2017) are disrupted in all *D. yakuba*-clade Wolbachia.
146 These Wolbachia also contain a set of *cin* loci, absent in *w*Mel, but very similar to those found in
147 *w*Pip (Beckmann *et al.* 2017), a strain that diverged 6–46 million years ago (mya) from *w*Yak
148 and *w*Mel (Werren *et al*. 1995; Meany *et al.* 2018). This is the first discovery of two sets of loci
149 implicated in CI being housed within the same prophage region. Several analyses implicate
150 Wolbachia-specific insertion sequence (IS) transposable elements, specifically ISWpi1, in the
151 horizontal transfer of these loci and surrounding regions between distantly related Wolbachia.
152 Horizontal movement of incompatibility factors between prophage regions of Wolbachia variants

153 adds another layer to what is already known about horizontal movement of prophages within and
154 between Wolbachia variants that themselves move horizontally between host species.
155

156 METHODS

157

158 Genomic data

159 The *D. yakuba*-clade isofemale lines included in our study were sampled over several years in
160 west Africa (Comeault *et al.* 2016; Turissini and Matute 2017; Cooper *et al.* 2017). Each line
161 within each species used in our analyses exhibits little nuclear introgression (< 1%) (Turissini
162 and Matute 2017), and no hybrids were included. Reads from *D. yakuba* (*N* = 56), *D. santomea*
163 (*N* = 11), and *D. teissieri* (*N* = 13) isofemale lines were obtained from the data archives of
164 Turissini and Matute (2017) and aligned to the *D. yakuba* nuclear and mitochondrial reference
165 genomes (Clark *et al*. 2007) with bwa 0.7.12 (Li and Durbin 2009), requiring alignment quality
166 scores of at least 50. Because many of the read archives were single end, all alignments were
167 completed using single-end mode for consistency.
168

169 **mtDNA:** Consensus mitochondrial sequences for each of the 80 lines were extracted with
170 samtools v. 1.3.1 and bcftools v 1.3.1 (Li 2011). Sequences for the 13 protein-coding genes were
171 extracted, based on their positions in the *D. yakuba* reference. We also extracted the 13 protein-
172 coding genes from additional unique *D. yakuba* (*N* = 28) and *D. santomea* (*N* = 15)
173 mitochondrial genomes (Llopart *et al.* 2014), and from the *D. melanogaster* reference (Hoskins
174 *et al.* 2015). Genes were aligned using MAFFT v. 7 and concatenated (Katoh and Standley
175 2013). Lines identical across all 13 genes were represented by one sequence for the phylogenetic
176 analyses.
177

178 **Wolbachia:** Reads from each of the 80 Turissini and Matute (2017) lines were aligned using
179 bwa 0.7.12 (Li and Durban 2009) to the *D. yakuba* reference genome (Clark *et al.* 2007)
180 combined with the *w*Mel reference genome (Wu *et al.* 2004). We calculated the average depth of
181 coverage across the *w*Mel genome. We considered lines with < 1× coverage uninfected and lines
182 with >10× coverage infected. No lines had between 1× and 10× coverage. To test our genomic
183 analyses of infection status, we used a polymerase chain reaction (PCR) assay on a subset of

184 lines. We extracted DNA using a "squish" buffer protocol (Gloor *et al*. 1993), and infection
185 status was determined using primers for the Wolbachia-specific *wsp* gene (Braig *et al.* 1998;
186 Baldo *et al.* 2006). We amplified a Drosophila-specific region of chromosome *2L* as a positive
187 control (Kern *et al.* 2015) (primers are listed in Supplemental Material, Table S1). For each run,
188 we also included a known Wolbachia-positive line and a water blank as controls.
189 To produce a draft Wolbachia "pseudoreference" genome for each species, we first
190 determined the isofemale line from each species that produced the greatest average coverage
191 depth over *w*Mel. We trimmed the reads with Sickle v. 1.33 (Joshi and Fass 2011) and assembled
192 with ABySS v. 2.0.2 (Jackman *et al.* 2017). *K* values of 51, 61…91 were tried. Scaffolds with
193 best nucleotide BLAST matches to known Wolbachia sequences with E-values less than 10^{-10}
194 were extracted as the draft Wolbachia assembly. For each species, the assembly with the highest
195 N50 and fewest scaffolds was kept as our Wolbachia pseudoreference genome for that host
196 species, denoted *w*Yak, *w*San, and *w*Tei (Table S2). To assess the quality of these three draft
197 assemblies, we used BUSCO v. 3.0.0 to search for homologs of the near-universal, single-copy
198 genes in the BUSCO proteobacteria database (Simão *et al.* 2015). For comparison, we followed
199 Conner *et al*. (2017) and performed the same search on the reference genomes for *w*Mel (Wu *et*
200 *al.* 2004), *w*Ri (Klasson *et al.* 2009), *w*Au (Sutton *et al.* 2014), *w*Ha and *w*No (Ellegaard *et al.*
201 2013) (Table S3).
202 Using these draft Wolbachia pseudoreference genomes, all other samples were aligned to
203 the *D. yakuba* reference (nuclear and mitochondrial) plus the species-specific Wolbachia draft
204 assembly with bwa 0.7.12 (Li and Durbin 2009). Consensus sequences were extracted with
205 samtools v. 1.3.1 and bcftools v. 1.3.1 (Li 2011).
206 To test whether the choice of Wolbachia pseudoreference influenced the Wolbachia draft
207 sequence obtained for each infected line, we arbitrarily selected three infected lines of each host
208 species and aligned reads from each of those lines independently to our *w*Yak, *w*San, and *w*Tei
209 pseudoreferences. This resulted in 27 alignments (9 per host species). Among the three
210 Wolbachia consensus sequences generated for each line, we assessed variation across all genes
211 used for downstream analyses. Irrespective of the pseudoreference used, we obtained identical
212 Wolbachia sequences. This indicates the robustness of our approach to generating population
213 samples of draft Wolbachia genomes, without having a high-quality Wolbachia reference

genome from any of the host species. Because Llopart *et al.* (2014) released only assembled mitochondrial genomes, we could not assess the Wolbachia infection status of their lines.

Loci for phylogenetic and comparative-rate analyses

Wolbachia genes: For each phylogenetic analysis of Wolbachia data, the draft genomes were annotated with Prokka v. 1.11 (Seemann 2014), which identifies homologs to known bacterial genes. To avoid pseudogenes and paralogs, we used only genes present in a single copy and with identical lengths in all of the sequences analyzed. Genes were identified as single copy if they uniquely matched a bacterial reference gene identified by Prokka v. 1.11. By requiring all homologs to have identical length in all of our draft Wolbachia genomes, we removed all loci with indels across any of our sequences.

Given that many loci accumulate indels over time, the number of loci included in our phylogenetic analyses depended on the number of strains included. We first estimated phylograms and chronograms for the "A group" Wolbachia from: the *D. yakuba*-clade, *w*Mel from *D. melanogaster* (Wu *et al*. 2004), *w*Inc that infects *Drosophila incompta* (Wallau *et al.* 2016), Wolbachia that infect Nomada bees (*w*NFe, *w*NPa, *w*NLeu, and *w*NFa; Gerth and Bleidorn 2016), and Wolbachia that infect *D. simulans* (*w*Ri, *w*Au and *w*Ha; Klasson *et al*. 2009; Sutton *et al*. 2014; Ellegaard *et al*. 2013). We also analyzed distantly related "B group" Wolbachia: *w*No from *D. simulans* (Ellegaard *et al*. 2013), *w*Pip_Pel from *Culex pipiens* (Klasson *et al*. 2008), and *w*AlbB from *Aedes albopictus* (Mavingui *et al.* 2012). To increase our phylogenetic resolution within the *D. yakuba* Wolbachia clade (by increasing the number of loci), we estimated a phylogram that included only *D. yakuba*-clade Wolbachia and *w*Mel. Our phylogram with both A- and B-group Wolbachia included 146 genes (containing 115,686 bp), our chronogram with only *D. yakuba*-clade Wolbachia included 678 genes (containing 695,118 bp), and our phylogram that included only *D. yakuba*-clade Wolbachia and *w*Mel included 643 genes (containing 644,586 bp),

To illustrate how divergence-time estimates vary with changing patterns of molecular divergence, we estimated two additional Wolbachia chronograms: one that considered only the *w*Mel reference plus the two most-diverged *w*Mel lines included in Richardson *et al.* (2012), and one that included these three *w*Mel variants and our *D. yakuba*-clade Wolbachia. Our chronogram with only *w*Mel Wolbachia included 692 genes (containing 709,599 bp), and our

245 chronogram that included *w*Mel plus *D. yakuba*-clade Wolbachia included 621 genes (containing
246 624,438 bp).
247
248 **mtDNA and nuclear genes from diverse Drosophila:** To assess variation in the relative rates
249 of divergence of mitochondrial and nuclear genes, we analyzed the canonical 12 Drosophila
250 genomes (Clark *et al.* 2007) plus *D. suzukii* (Chiu *et al*. 2013). We excluded *D. sechellia* and *D.*
251 *persimilis* which show evidence of introgression with *D. simulans* and *D. pseudoobscura*,
252 respectively (Kulathinal *et al.* 2009; Schrider *et al.* 2018; Brand *et al.* 2013). Coding sequences
253 for 20 nuclear genes used in the analyses of Turelli *et al.* (2018) (*aconitase, aldolase, bicoid,*
254 *ebony, enolase, esc, g6pdh, glyp, glys, ninaE, pepck, pgi, pgm, pic, ptc, tpi, transaldolase, white,*
255 *wingless*, and *yellow*) were obtained from FlyBase for each species. The genes were aligned with
256 MAFFT v. 7 (Katoh and Standley 2013). Coding sequences for the 13 protein-coding
257 mitochondrial genes in the inbred reference strains were also obtained from FlyBase for each
258 species and were aligned with MAFFT v. 7 and concatenated.
259
260 **Phylogenetic analyses**
261 All of our analyses used RevBayes v. 1.0.9 (Hohna *et al.* 2016), following the procedures of
262 Turelli *et al*. (2018). For completeness, we summarize those methods below. For additional
263 details on the priors and their justifications, consult Turelli *et al*. (2018). Four independent runs
264 were performed for each analysis, and in all cases runs agreed with one another. Nodes with
265 posterior probability less than 0.95 were collapsed into polytomies.
266
267 **Wolbachia phylograms:** We estimated a phylogram for A- and B-group Wolbachia and for only
268 *D. yakuba*-clade and *w*Mel Wolbachia using the same methodology as Turelli *et al.* (2018). We
269 used a GTR + Γ model with four rate categories, partitioning by codon position. Each partition
270 had an independent rate multiplier with prior Γ(1,1) (i.e., Exp(1)), as well as stationary
271 frequencies and exchangeability rates drawn from flat, symmetrical Dirichlet distributions (i.e.,
272 Dirichlet(1,1,1...). The model used a uniform prior over all possible topologies. Branch lengths
273 were drawn from a flat, symmetrical Dirichlet distribution, thus they summed to 1. Since the
274 expected number of substitutions along a branch equals the branch length times the rate

275 multiplier, the expected number of substitutions across the entire tree for a partition is equal to
276 the partition's rate multiplier.
277
278 **Wolbachia chronograms:** We first created a relaxed-clock relative chronogram with the root
279 age fixed to 1 using the GTR + Γ model, partitioned by codon position, using the same birth-
280 death prior as Turelli *et al.* (2018). Each partition had an independent rate multiplier with prior
281 Gamma(1,1), as well as stationary frequencies and exchangeability rates drawn from flat,
282 symmetrical Dirichlet distributions. The branch-rate prior for each branch was
283 Γ(2,2), normalized to a mean of 1 across all branches (Table S4). We also tried a strict-clock tree
284 and a relaxed-clock tree with branch-rate prior Γ(7,7), which produced no significant differences.
285 We used the scaled distribution $\Gamma(7,7) \times 6.87 \times 10^{-9}$ to model substitutions per third-position site
286 per year. This transforms the relative chronogram into an absolute chronogram. This scaled
287 distribution was chosen to replicate the upper and lower credible intervals of the posterior
288 distribution estimated by Richardson *et al*. (2012), assuming 10 Drosophila generations per year,
289 normalized by their median substitution-rate estimate. Branch lengths in absolute time were
290 calculated as the relative branch length times the third position rate multiplier divided by the
291 substitutions per third position site per year estimate above.
292 We illustrate how divergence-time estimates depend on changing patterns of Wolbachia
293 molecular evolution observed over different time scales, specifically the relative rates of third-
294 site substitutions versus first- and second-site substitutions. We compare divergence times
295 estimated for variants within *melanogaster*-group host species, with a time scale of only
296 hundreds or thousands of years, with the same divergence times estimated when more distantly
297 related Wolbachia, with divergence times over tens of thousands of years, are included in the
298 analyses. We present three separate analyses: one using only *D. yakuba*-clade Wolbachia
299 variants, a second using only *w*Mel variants analyzed by Richardson *et al.* (2012), and a third
300 simultaneously analyzing both sets of data.
301
302 ***Drosophila* phylogeny for 11 references species:** We estimated a phylogram from our 20
303 nuclear loci using the GTR + Γ model, with partitioning by gene and codon position. The model
304 was identical to the Wolbachia phylogram model above except for the partitioning (there are too
305 few Wolbachia substitutions to justify partitioning by gene).

306 **mtDNA phylogeny for the 11 reference species**: We estimated a phylogram from the 13
307 protein-coding mitochondrial loci using the GTR + Γ model, with partitioning only by codon
308 position. The model was identical to the Wolbachia phylogram model above.
309
310 **mtDNA chronogram for the *D. yakuba* clade:** We estimated a relative chronogram for the *D.*
311 *yakuba*-clade mitochondria from the 13 protein-coding loci using the GTR + Γ model,
312 partitioning only by codon position. To test the sensitivity of our results to priors, we ran a strict
313 clock, a relaxed clock with a Γ(7,7) branch-rate prior, and a relaxed clock with a Γ(2,2) branch-
314 rate prior, corresponding to increasing levels of substitution-rate variation across branches. The
315 model was identical to the Wolbachia chronogram model above, except that we did not
316 transform it into an absolute chronogram.
317
318 **Ratios of divergence rates for mtDNA versus nuclear loci:** To quantify ratios of mtDNA to
319 nuclear substitution rates, we estimated relative substitution rates for host nuclear genes versus
320 mtDNA using the GTR + Γ model. The (unrooted) topology was fixed to the consensus topology
321 for the nuclear and mitochondrial data. The data from 20 nuclear loci were partitioned by locus
322 and codon position, the mtDNA data were partitioned only by codon position. All partitions
323 shared the same topology, but the nuclear partitions were allowed to have branch lengths
324 different from the mtDNA partitions. The sum of the branch lengths for each partition was scaled
325 to 1. Assuming concurrent nuclear-mtDNA divergence (because we used only species showing
326 no evidence of introgression), we imposed the same absolute ages for all nodes of the nuclear
327 and mtDNA chronograms. For each nuclear locus, the third-position rate-ratio for the mtDNA
328 versus nuclear genomes was calculated as: (mitochondrial branch length × mitochondrial 3rd
329 position rate multiplier)/(nuclear branch length × nuclear 3rd position rate multiplier). We
330 summarized the relative rates of mtDNA versus nuclear substitutions along each branch using
331 the arithmetic average of the 20 ratios obtained from the individual nuclear loci.
332
333 **Introgressive versus horizontal Wolbachia transfer–concordance of phylograms:** Horizontal
334 transfer of Wolbachia was initially detected by discordance between Wolbachia and host
335 phylogenies (O'Neill *et al*. 1992). When considering samples within species and closely related
336 species, a comparison of mtDNA and Wolbachia phylogenies provides a natural test for

horizontal transmission. We first look for significant discrepancies between strongly supported nodes in the mtDNA versus Wolbachia phylogenies. To test for less obvious differences, we follow Richardson *et al*. (2012) and compute Bayes Factors to assess the support for models that assume that mtDNA versus Wolbachia follow the same topology versus distinct phylogenies. For these calculations, we omit san-Quija37 due to its clear discordance.

We calculated the marginal likelihood in RevBayes for two models; one with a shared mtDNA and Wolbachia phylogeny, another with independent topologies. The mtDNA and Wolbachia were each partitioned by codon position, for a total of six partitions. In the shared phylogeny model, all six share the same topology; in the independent model, the mtDNA and Wolbachia have separate topologies. All priors were the same as the Wolbachia phylogram model. Lines that were identical across the mtDNA and Wolbachia were collapsed into one sample. We ran two independent replicates of each model with 50 stepping stones per run. The Bayes factor is computed as the difference between the marginal likelihoods of each model.

Introgressive versus horizontal Wolbachia transfer–relative rates: Within species, the phylogenies of mtDNA and Wolbachia are fully resolved. Thus we consider an alternative approach for distinguishing between introgressive versus horizontal transfer that depends on estimating divergence times for Wolbachia genomes and host mtDNA. This is complicated by variation in the relative rates of mtDNA versus Wolbachia divergence and by systematic changes over time in the relative rates of substitutions at the three codon positions for Wolbachia. Following the procedures in Turelli *et al.* (2018), we estimated the mtDNA-to-Wolbachia third-position substitution-rate ratio for each branch in the *D. yakuba* clade. For each analysis, we computed the marginal likelihood of the model where all branches shared the same ratio and the same model except allowing different ratios on each branch. We then calculated Bayes factors (i.e. differences in the log of the marginal likelihoods) to determine which model was favored. We repeated these analyses after including three *w*Mel-infected *D. melanogaster* lines with the *D. yakuba*-clade dataset. Because *D. yakuba*-clade species and *D. melanogaster* do not produce fertile hybrids (Sánchez and Santamaria 1997; Turissini *et al.* 2017), introgressive transfer of Wolbachia is not possible between these species. Hence, including *w*Mel provides a control for our proposed test of introgressive transfer. For these controls, we calculated the mitochondrial

367 versus Wolbachia substitution-rate ratio for the interspecific branches separately, regardless of
368 whether the shared rate-ratio model was favored.
369
370 **Wolbachia loci associated with CI**
371 Wolbachia that infect all three *D. yakuba*-clade hosts cause weak intra-and interspecific CI
372 (Cooper *et al.* 2017), but the genetic basis of CI in this clade remains unknown. We used protein
373 BLASTN to search for *cif* homologs in each of our *D. yakuba*-clade Wolbachia assemblies,
374 querying *cidA-cidB* variants and the paralogous pair *cinA-cinB*$^{w\mathrm{Ri}}$ (wRi_006720 and
375 wRi_006710) found in *w*Ri and some *w*Ri-like Wolbachia (Beckmann and Fallon 2013; Turelli
376 *et al*. 2018), and the *cinA-cinB*$^{w\mathrm{Pip}}$ pair (Beckmann and Fallon 2013; Beckmann *et al*. 2017;
377 LePage *et al*. 2017; Lindsey *et al*. 2018). We identified both *cid* and *cin* homologs, but we found
378 no close matches to the *cinA-cinB*$^{w\mathrm{Ri}}$ pair (denoted Type II *cif* loci by Lepage *et al*. 2017) in any
379 of our genomes. For all of the samples, we extracted consensus sequences from our assemblies
380 and alignments for the *cidA-cidB*$^{w\mathrm{Yak\text{-}clade}}$ and *cinA-cinB*$^{w\mathrm{Yak\text{-}clade}}$ gene pairs. The genes were
381 aligned with MAFFT v. 7 (Katoh and Standley 2013). We examined variation in these regions
382 relative to *cidA-cidB*$^{w\mathrm{Mel}}$ and *cinA-cinB*$^{w\mathrm{Pip}}$ respectively.
383 Because we unexpectedly identified homologs of *cinA-cinB* in all *w*Yak-clade Wolbachia,
384 we took additional measures to understand the origin and placement of these loci in the genomes.
385 The *cinA-cinB*$^{w\mathrm{Yak\text{-}clade}}$ open reading frames are located on a ~11,500 bp scaffold in the
386 fragmented *w*Yak assembly (*w*Yak scaffold "702380" from ABySS output—we use quotes to
387 indicate names assigned by ABySS). The first ~4,000 bp of this scaffold contain *cinA* and *cinB*
388 genes that have ~97% identity with those in *w*Pip. We performed a BLAST search using all
389 contigs in the *w*Yak assembly as queries against the *w*Mel genome (Camacho *et al*. 2009). The
390 *cinA-cinB*$^{w\mathrm{Yak\text{-}clade}}$ scaffold was placed on the *w*Mel genome (~617,000–623,000) adjacent to and
391 downstream of the *w*Yak contig containing *cidA-cidB*$^{w\mathrm{Yak\text{-}clade}}$ loci (*w*Yak contig "187383").
392 However, only ~7,000 bp of the 11,500 bp of the scaffold containing *cinA-cinB*$^{w\mathrm{Yak\text{-}clade}}$ align to
393 this region of *w*Mel. The ~4,000 bp sequence that contains the ORFs for *cinA-cinB*$^{w\mathrm{Yak\text{-}clade}}$ has
394 no significant hit against the *w*Mel genome.
395 To verify the placement of the scaffold containing *cinA-cinB*$^{w\mathrm{Yak\text{-}clade}}$, we first performed
396 targeted assembly using the *cinA-cinB*$^{w\mathrm{Yak\text{-}clade}}$ contig and the adjacent contigs as references
397 (Hunter *et al*. 2015). Iterative targeted assembly often extends assembled scaffolds (Langmead

398 and Salzberg 2012). Reads are mapped to the target of interest, and the mapped reads are then
399 assembled using SPADES (Bankevich *et al*. 2012; Nurk *et al*. 2013). The newly assembled
400 scaffolds serve as the target in a subsequent round of mapping. The procedure is repeated until
401 no new reads are recruited to the assembled scaffold from the previous round. The extended
402 scaffolds enabled us to merge flanking *w*Yak contigs downstream of the *cinA-cinB*$^{w\text{Yak-clade}}$
403 scaffold but failed to connect the fragment with the *cidA-cidB*$^{w\text{Yak-clade}}$ contig. We used PCR to
404 amplify the intervening region (Sanger *et al.* 1977) (Supplemental Information, Table S1), and
405 Sanger sequencing to evaluate this product.
406 Subsequent mapping of paired-end reads to the merged scaffolds confirmed the correct
407 order and orientation of the contigs containing the *cidA-cidB*$^{w\text{Yak-clade}}$ and *cinA-cinB*$^{w\text{Yak-clade}}$
408 genes. Observed, uncorrected, pairwise distances between focal regions in *w*Yak, *w*Mel, and
409 *w*Pip and between focal regions in *w*Yak, *w*Pip, *w*AlbB, and *w*Nleu were calculated using a
410 sliding window (window size = 200 bp, step size = 25 bp).
411

412 # RESULTS

413

414 **Mitochondrial introgression**

415 Previous analyses suggest that *D. yakuba* and *D. santomea* carry very similar mtDNA due to
416 ongoing hybridization and introgression (Lachaise *et al*. 2000; Bachtrog *et al.* 2006; Llopart *et*
417 *al.* 2014). However, our mtDNA relative chronogram (Figure 1), based on mitochondrial whole-
418 proteome data sampled from throughout the ranges of all three species, supports three mtDNA
419 clades that largely agree with the nuclear topology—*D. teissieri* mtDNA sequences are outgroup
420 to sister *D. yakuba* and *D. santomea* mtDNA (Figure 1). Consistent with introgression, 12 out of
421 26 *D. santomea* isofemale lines have *D. yakuba*-like mtDNA (indicated by blue branches and
422 letters in Figure 1). Yet, our samples showed no evidence of *D. santomea* mtDNA introgressed
423 into *D. yakuba* and no mtDNA introgression involving *D. teissieri*. The 12 *D. santomea*
424 isofemales with introgressed *D. yakuba*-like mtDNA were sampled from both within (N = 8) and
425 outside (N = 4) the well-described Pico de São Tomé hybrid zone (Llopart et al. 2014; Turissini
426 and Matute 2017). The *D. santomea* hybrid-zone samples are indicated by "HZ" in Figure 1 in
427 blue, and they are found throughout the clade that includes all *D. yakuba* mtDNA. This suggests

428 that hybridization occurs in other areas of the island or that introgressed *D. santomea* genotypes
429 migrate outside of the hybrid zone.
430 Previous analysis of nuclear genomes (Turissini and Matute 2017) also found more
431 introgression from *D. yakuba* into *D. santomea* when populations of *D. yakuba* from near the
432 Gulf of Guinea, Cameroon, and Kenya were included; however, excluding Cameroon and Kenya
433 indicated similar amounts of introgression in each direction. Matings between *D.*
434 *santomea* females and *D. yakuba* males are rare relative to the reciprocal cross (Coyne *et al*.
435 2002, Matute 2010). When these matings do occur, F1 females produce fewer progeny than do
436 F1 females produced by *D. yakuba* females and *D. santomea* males (Matute and Coyne 2010).
437 F1 hybrids produced by *D. santomea* females also have a shortened lifespan (Matute and Coyne
438 2010), due to copulatory wounds inflicted by *D. yakuba* males during mating (Kamimura 2012).
439 These observations are consistent with our finding of preferential introgression of *D.*
440 *yakuba* mitochondria into *D. santomea* backgrounds.
441
442 **Wolbachia frequencies and draft genomes**
443 As expected, we found Wolbachia in all three *D. yakuba*-clade species sampled by Turissini and
444 Matute (2017). For *D. yakuba,* 21 of 56 lines were infected, yielding an infection frequency of p
445 = 0.36, with 95% binomial confidence interval (0.25, 0.51). For *D. santomea*, 10 of 11 were
446 infected, p = 0.91 (0.59, 1.0); and for *D. teissieri*, 11 of 13 were infected, p = 0.85 (0.55, 0.98).
447 Additional frequency estimates are reported in Cooper *et al*. (2017), which found that Wolbachia
448 frequencies vary through time and space in west Africa. All PCR tests for Wolbachia matched
449 our coverage-based genomic analyses of infection status.
450 The average read depth over the *w*Mel reference sequence was 1,940 for our *w*Yak
451 pseudoreference genome (yak-CY17C), 40 for *w*San (san-Quija630.39), and 489 for *w*Tei (teis-
452 cascade_4_2). These pseudoreference genomes were included in the phylogram that includes A-
453 and B-group Wolbachia (Figure 2), and they cluster with the Wolbachia from their respective
454 hosts (Figure 3), as expected if introgression is rare. In general, the *w*San data are lower quality
455 due to the relatively small size of the *D. santomea* libraries (Turissini and Matute 2017). The
456 scaffold count, N50, and total assembly sizes are reported in Supplementary Table S2.
457
458

459

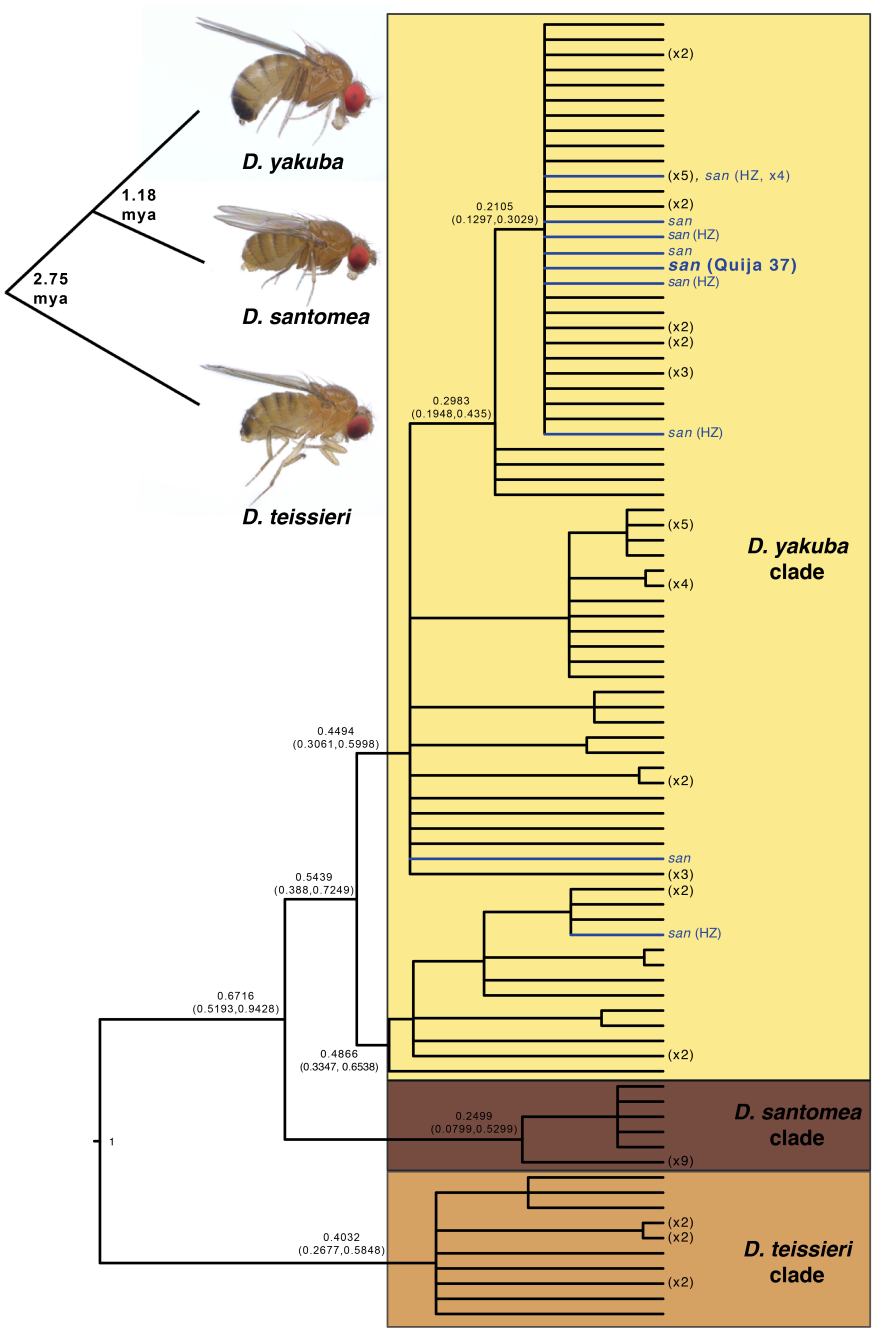

Figure 1 Bayesian relative chronogram for the *D. yakuba* clade mtDNA. The root age was fixed to 1. Nodes with posterior probability less than 0.95 were collapsed into polytomies, numbers at the tips show the number of genotypes collapsed. Node labels are the point estimates and 95% credible intervals for the relative ages. Tips and branches showing *D. santomea* individuals with *D. yakuba* mtDNA are colored blue, those sampled from the Pico de São Tomé hybrid zone are denoted by HZ. *D. santomea* line Quija 37 showing evidence of horizontal or paternal Wolbachia transmission is in bold blue. The chronogram was estimated from the thirteen protein-coding mitochondrial genes. A representation of host nuclear relationships with photos of host females at the tips is inset, with the divergence times estimated by Turissini and Matute (2017) superimposed.

460 **Wolbachia phylograms, chronograms and relative rates**

461 Our phylogram that includes A- and B-group Wolbachia places the three weak-CI-causing *D.*
462 *yakuba*-clade Wolbachia together and sister to *w*Mel in the A group (Figure 2A). In contrast, *D.*
463 *simulans* carries diverse Wolbachia that do (*w*Ri, *w*Ha, *w*No) and do not (*w*Au) cause CI,
464 spanning Wolbachia groups A (*w*Ri, *w*Ha, *w*Au) and B (*w*No). All *w*San, *w*Yak, and *w*Tei
465 Wolbachia variants included in our analysis are identical at the MLST loci often used to type
466 Wolbachia variants (Cooper *et al.* 2017; Baldo *et al.* 2006). Due to the relatively small number
467 of genes that meet our inclusion criteria (146 genes across 115,686 bp), the resulting phylogram
468 does not resolve relationships among *w*San, *w*Yak, and *w*Tei (Figure 2A). Our phylogram that
469 includes only *D. yakuba*-clade and *w*Mel Wolbachia (based on 643 genes across 644,586 bp)
470 resolves these relationships, placing *w*Tei variants outgroup to sister *w*San and *w*Yak in distinct
471 clades (Figure 2B). We observe 0.0039% third-position pairwise differences between *w*Tei and
472 *w*Yak and between *w*Tei and *w*San, and 0.0017% between *w*San and *w*Yak. The third-position
473 pairwise differences between any of the *D. yakuba*-clade Wolbachia and *w*Mel is 0.11%. For
474 reference, the third-position pairwise difference between *w*Yak and *w*Ri is 2.9%, whereas *w*Ri
475 and *w*Suz differ by only 0.014% (Turelli *et al.* 2018).

476 Our absolute chronogram that includes *w*Mel and *D. yakuba*-clade Wolbachia also indicates
477 three monophyletic groups (Figure 3), whose topology—(*w*Tei, (*w*Yak, *w*San))—agrees with the
478 nuclear topology of the hosts (Turissini and Matute 2017). However, one *D. santomea* line, *D.*
479 *santomea* Quija 37, sampled from the south side of São Tomé (Matute 2010), carries Wolbachia
480 identical to the other eight *w*San samples across 695,118 bp, yet its mtDNA is nested in the *D.*
481 *yakuba* mtDNA cluster (bold blue in Figure 1). This single example of Wolbachia-mtDNA
482 phylogenetic discordance can be explained by either horizontal Wolbachia transfer or paternal
483 transfer of Wolbachia or mtDNA. While rare, paternal Wolbachia transmission has been
484 observed in other Drosophila (Hoffmann and Turelli 1988; Turelli and Hoffmann 1995), as has
485 paternal transmission of mtDNA (Kondo *et al.* 1990). The only other Wolbachia-screened *D.*
486 *santomea* isofemale line with introgressed mtDNA was Wolbachia uninfected.

487 In Figure 3, we present alternative estimates of the divergence times for Wolbachia within
488 and between *D. yakuba*-clade species, within *D. melanogaster*, and between the *D. yakuba* clade
489 and *D. melanogaster.* The estimates from the chronogram that includes all of the indicated
490 Wolbachia variants are in bold. To demonstrate how these time estimates depend on the

491 sequences included in the analyses (through estimates of relative divergence for each codon
492 position), estimates that included only *D. yakuba*-clade Wolbachia or only *w*Mel variants are
493 superimposed and not in bold. These estimates were all obtained using a relaxed-clock analysis
494 with a Γ(2,2) branch-rate prior (results using Γ(7,7) and a strict clock are mentioned below).
495 When considering only the *D. yakuba*-clade Wolbachia (Figure 3), we estimate the root at 2,510
496 years (95% credible interval = 773 to 5,867 years) [our point estimate is 2,642 and credible
497 interval (781, 5,904) using Γ(7,7)] and the *w*Yak-*w*San split at 1,556 years (95% credible interval
498 = 411 to 3,662 years) [point estimate 1,545 and credible interval (401, 3,680) with Γ(7,7)]. The
499 *w*Mel chronogram estimates the most recent common ancestor (MRCA) of *w*Mel at 4,890 years
500 ago (95% credible interval = 1,258 to 11,392 years) [point estimate 4955 and credible interval
501 (1363, 11093) with Γ(7,7)]. Our *w*Mel result is consistent with the Richardson *et al.* (2012)
502 estimate of 8,008 years (95% credible interval (3,263, 13,998). As noted, the alternative branch-
503 rate priors, Γ(2,2) versus Γ(7,7), have little effect on these divergence-times estimates.
504

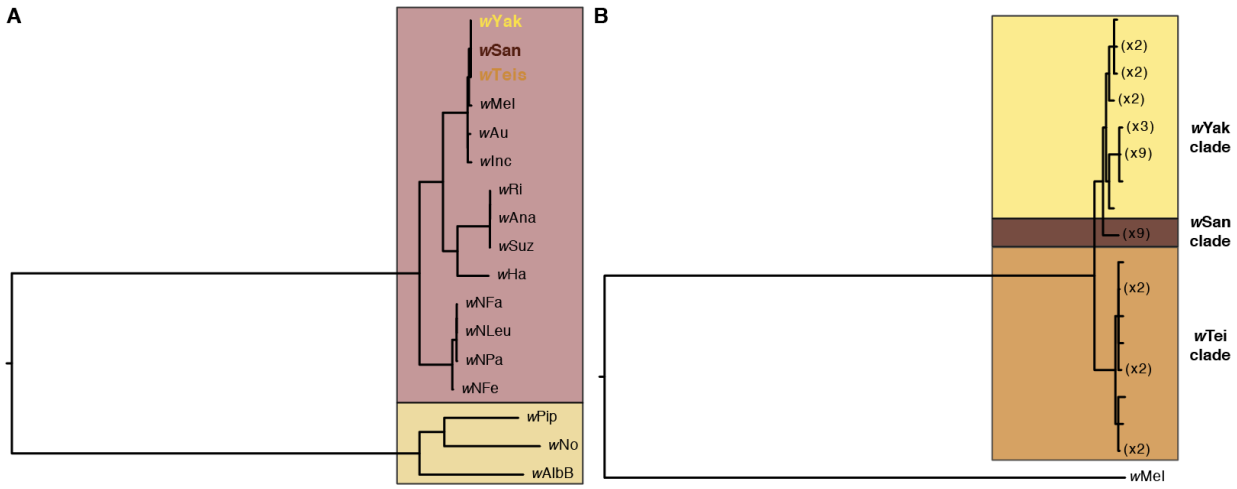

Figure 2 A) Bayesian phylogram placing the *D. yakuba*-clade Wolbachia in the context of other A-, and also B-group, Wolbachia. *w*Pip, wNo, and wAlbB belong to Wolbachia group B, while *D. yakuba* clade-Wolbachia, *w*Mel, *w*Au, *w*Inc, *w*Ri, *w*Ana, *w*Suz, *w*Ha, and Nomada Wolbachia belong to group A. The phylogram was estimated with 146 single-copy genes of identical length in each genome, spanning 115,686 base pairs. The limited number of genes meeting our criteria precluded resolving relationships of *D. yakuba*-clade Wolbachia. B) Bayesian phylogram of *w*Yak, *w*San, *w*Tei, and outgroup *w*Mel. The phylogram was estimated with 643 single-copy genes of identical length in each genome, spanning 644,586 bp. These more extensive data provide increased phylogenetic resolution of the *D. yakuba*-clade Wolbachia. For both phylograms, nodes with posterior probability less than 0.95 were collapsed into polytomies.

505

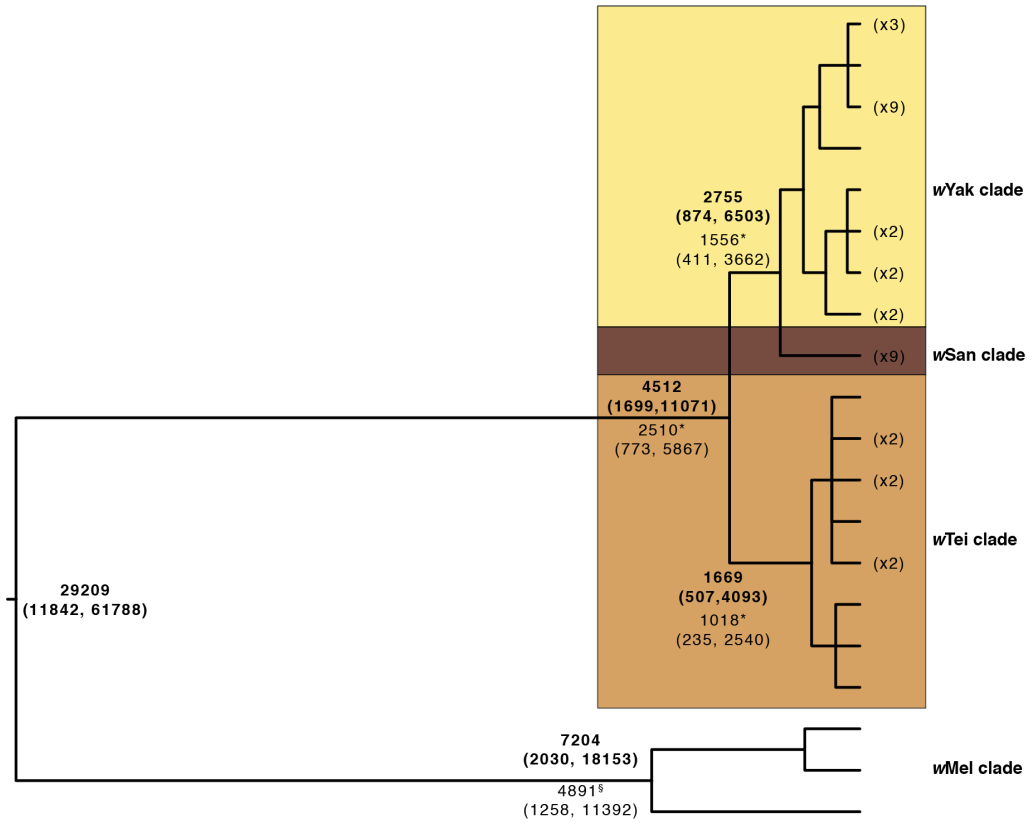

Figure 3 Bayesian chronogram for the *D. yakuba*-clade Wolbachia plus three *w*Mel genomes with absolute age estimates based on the calibration in Richardson *et al*. (2012) in bold. The chronogram was estimated with 621 single-copy genes of identical length in all of the genomes, spanning 624,438 bp. Nodes with posterior probability less than 0.95 were collapsed into polytomies. Node labels are the point estimates and 95% credible intervals for the ages. We also constructed two other chronograms: one that included only *D. yakuba*-clade Wolbachia and one that included only *w*Mel Wolbachia. The divergence estimates for the chronogram including only *D. yakuba*-clade Wolbachia (*) and only *w*Mel Wolbachia (§) are superimposed to illustrate that the inclusion of both groups in the chronogram alters the absolute divergence estimates within each group. Our chronogram with only *D. yakuba*-clade Wolbachia included 678 genes (containing 695,118 bp), and our chronogram including only *w*Mel Wolbachia included 692 genes (709,599 bp).

506 In contrast, Figure 3 shows that including more divergent variants significantly influences our
507 time estimates. When we analyze the *D. yakuba*-clade and *w*Mel Wolbachia, using the same
508 Γ(2,2) relaxed-clock model, we estimate the root of the *D. yakuba*-clade Wolbachia at 4,512
509 years (95% credible interval = 1,432 to 9,024 years) and the MRCA of *w*Mel at 7,204 years
510 (95% credible interval = 2,511 to 18,739 years), rather than 2,510 and 4,891, respectively.
511 Supplementary Table S4 shows how estimates of the relative amounts of the divergence for the

512 three codon positions vary across the data sets. The underlying model for divergence times
513 assumes constant relative rates at the three positions. Hence, changes in relative rates affect
514 divergence-time estimates.
515 As noted above, our estimates of the time of the MRCA for Wolbachia within the *D. yakuba*
516 clade or within *D. melanogaster* are relatively insensitive to whether the branch-rate prior is
517 $\Gamma(2,2)$ or $\Gamma(7,7)$. The estimated divergence time between *w*Mel and the *D. yakuba*-clade
518 Wolbachia is less robust. Using $\Gamma(2,2)$, we estimate this split at 29,209 years (95% credible
519 interval: 11,842 to 61,788 years); whereas $\Gamma(7,7)$ produces 49,585 years (17,452 to 106,310
520 years). (A strict clock analysis, corresponding to $\Gamma(n,n)$ as $n \to \infty$, produces 72,612, with 95%
521 credible interval (23,412, 132,276)). An alternative point estimate can be obtained from pairwise
522 differences. Averaging over the *D. yakuba*-clade sequences, we observe an average third-
523 position difference of 0.107% between the *D. yakuba*-clade Wolbachia and *w*Mel, or 0.0535%
524 from tip to root. The "short term evolutionary rate" of divergence within *w*Mel estimated by
525 Richardson *et al.* (2012) produces a point estimate of 78,000 years, which overlaps with the
526 credible interval of our relaxed-tree estimate with a $\Gamma(7,7)$ branch-rate prior. This molecular-
527 clock estimate closely agrees with our strict-clock chronogram analysis. Given that these
528 estimates are much shorter than the divergence time between their reproductively isolated hosts,
529 horizontal Wolbachia transmission must have occurred. As discussed below, the horizontal
530 transmission probably involved intermediate hosts.
531
532 ***Drosophila*-wide mtDNA to nuclear rate-ratio analysis:** The topologies of the nuclear and
533 mtDNA trees for the 11 reference Drosophila species are completely concordant (Figure 4) and
534 agree with the neighbor-joining result presented in Clark *et al*. (2007). We find the average ratio
535 of mtDNA to nuclear substitution rates across branches vary from about 1× to 10×, consistent
536 with the results of Havird and Sloan (2016) across Diptera.
537
538 **Comparison of mtDNA versus Wolbachia–concordance:** The discordance between the
539 placement of the mtDNA of *D. santomea* line Quija 37, which has mtDNA belonging to the
540 clade associated *D. yakuba* and its Wolbachia, which is clearly *w*San, demonstrates horizontal or
541 paternal transmission. Apart from this, we find no nodes in either the mtDNA or Wolbachia tree
542 with posterior probability at least 0.95 that is inconsistent with a strongly supported node (i.e.,

543 posterior probability at least 0.95) in the other tree. For a more refined test of concordance, we
544 removed the Quija 37 line from our mtDNA and Wolbachia data and computed Bayes Factors.
545 The shared topology model was favored by a Bayes factor of e^{55}, indicating strong support for
546 concordance of the mtDNA and Wolbachia phylogenies.
547

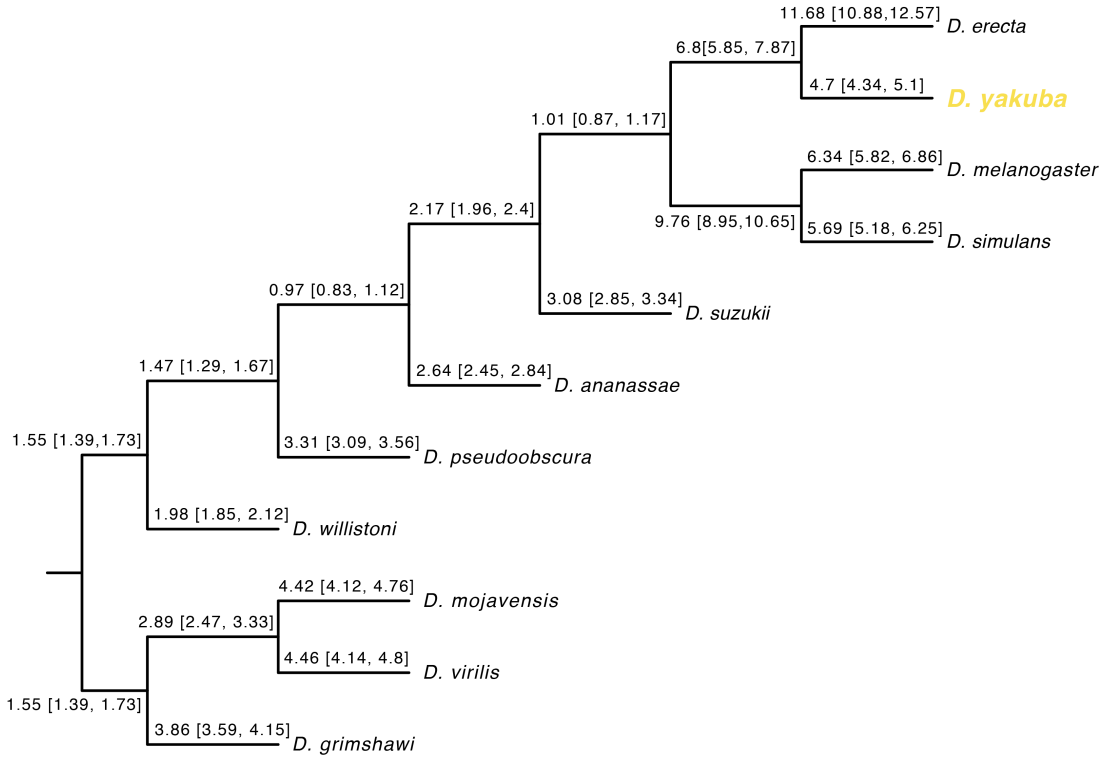

Figure 4 A Bayesian phylogram, with branch lengths fixed to 1, displaying relative rate variation for mtDNA versus nuclear loci. Branches are labeled with quartiles of the estimated third position mtDNA to third position nuclear substitution rate. The nuclear substitution rate was taken as the arithmetic mean third-position rate over the 20 loci used. The mtDNA and nuclear phylograms are topologically identical when run separately.

548 **Comparison of mtDNA versus Wolbachia–relative rates:** Following the methods of Turelli *et*
549 *al.* (2018) to look for different mtDNA versus Wolbachia branch lengths, we also found no
550 evidence of variation in the ratio of mtDNA:Wolbachia substitution rates within the *D. yakuba*
551 clade (median 3rd position ratio = 223, quartiles = 186 and 268); the model with all branches
552 sharing the same ratio is favored over the model where each branch has its own ratio by a Bayes
553 factor of e^{7}. For comparison, Turelli *et al.* (2018) found a median ratio of 566 within *D. suzukii*
554 and 406 within *w*Ri-infected *D. ananassae*-subgroup species. These differences in relative rates

555 are broadly compatible with the tenfold variance in relative rates noted above for nuclear versus
556 mitochondrial divergence across Drosophila species. Hence, our data seem compatible with
557 Wolbachia transmission via introgression within the *D. yakuba* clade.
558 Although this result seems to strongly support purely maternal transmission of Wolbachia
559 within species and introgressive transfer between species (with more recent introgression
560 between the sister species *D. yakuba* and *D. santomea*), this interpretation is severely weakened
561 by our "control" analysis that includes *w*Mel-infected *D. melanogaster* (for which introgression
562 is impossible). Including the *D. melanogaster* data, the constant-ratio model was still favored
563 over variable ratios by a Bayes factor of e^{54} (median 3rd position ratio = 297, quartiles = 276 and
564 323). We discuss the implications of this anomalous result below. The key observation is that
565 significantly declining rates of substitution for Wolbachia (and mtDNA, see Ho *et al.* 2005) over
566 time, together with heterogeneity of relative rates (as illustrated in Figure 4), limit the power of
567 relative substitution ratios to differentiate between introgression and horizontal transmission.
568
569 **Analysis of CI factors**
570 In contrast to *w*Mel, which contains only the *cidA-cidB* gene pair (LePage et al. 2017), the *D.*
571 *yakuba*-clade Wolbachia, which also belong to Wolbachia group A, have both *cidA-cidB* and a
572 *cinA-cinB* pair homologous to CI loci originally identified in group-B *w*Pip (Beckmann *et al.*
573 2017). The A and B Wolbachia groups diverged about 6-46 mya (Werren *et al.* 1995; Meany *et*
574 *al.* 2018). Among our *w*Yak-clade sequences, there are no single-nucleotide variants within any
575 of the *cif* loci. The $cidB^{w\text{Yak-clade}}$ locus has an inversion from amino acids 37–103 relative to the
576 same region in $cidB^{w\text{Mel}}$ in every *w*Yak-clade Wolbachia variant we analyzed. This introduces
577 several stop codons, which might render this gene nonfunctional. On the other hand, RNA
578 polymerase should still transcribe a complete polycistronic transcript. Therefore translation of an
579 N-terminally truncated $\text{CidB}^{w\text{Yak-clade}}$ protein cannot be ruled out. The sequence differences
580 between $cinA^{w\text{Yak-clade}}$ and $cinB^{w\text{Yak-clade}}$ regions compared to $cinA^{w\text{Pip}}$ and $cinB^{w\text{Pip}}$ homologs are
581 1.35% and 0.59%, respectively. In contrast, the average difference between *w*Yak and *w*Pip
582 genomes across all 194 genes (161,655 bp) present in single copy and with identical lengths in
583 both genomes is 11.60%. Conversely, outside of the prophage regions, *w*Mel and *w*Yak differ by
584 about 1%. This is consistent with data that indicate WO phage regions have a different

evolutionary history than the bulk of the Wolbachia genome (LePage *et al*. 2017; Lindsey *et al*. 2018).

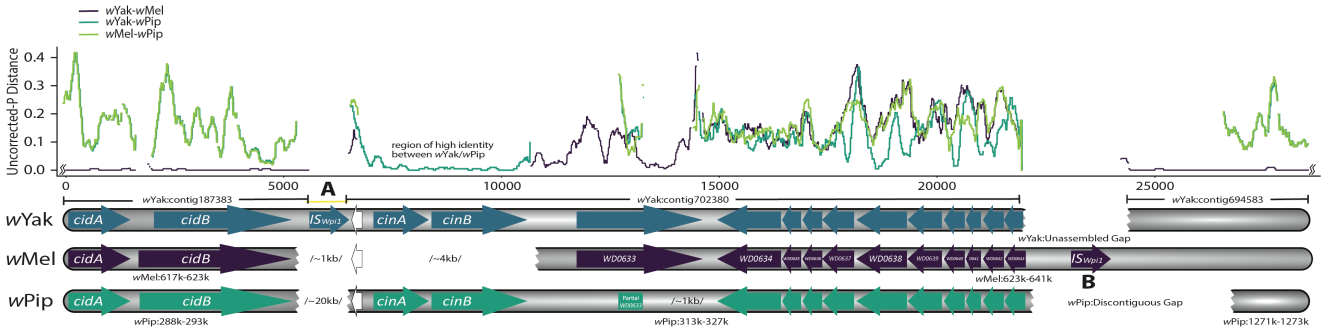

Figure 5 The gene structure and observed, uncorrected, pairwise distances among *w*Yak, *w*Mel, and *w*Pip Wolbachia spanning the region of the WO prophage island where the hypothesized transfer occurred between the ISWpi1 element "A" present in *w*Yak and the ISWpi1 sequence labeled "B" in *w*Mel. Over the region shown, *w*Yak, *w*San and *w*Tei are identical. The difference between *w*Yak and *w*Mel is less than 1% between the contigs (*w*Yak "187383" and *w*Yak "694583") flanking the region containing the *cinA-cinB* homologs. These homologs assembled on the first ~4,000 bp of the contig *w*Yak "702380", but they do not appear to exist in *w*Mel. A BLAST search against the NCBI nucleotide collection using contig *w*Yak "702380" as a query indicated the *cinA-cinB* homologs share high identity with group-B *w*Pip Wolbachia (~97%). The remainder of the *w*Yak contig is approximately equally divergent from both *w*Mel and *w*Pip, consistent with the entire region being transferred from an unknown Wolbachia variant that had acquired the *cinA-cinB* homologs. Under this scenario, ISWpi1 elements mediate the transfer to *w*Yak via excision of elements "A" and "B" along with the intervening DNA from the donor and subsequent homologous recombination between IS sequences "A" and "B" and WO phage island genes in *w*Yak. The ISWpi1 element "A" present in *w*Yak, but not in either *w*Mel or *w*Pip, was spanned using Sanger sequencing with primers anchored on unique sequence in the flanking contigs (Supplementary Table 2). The ISWpi1 sequence labeled "B" is assembled in *w*Mel and corresponds to *w*Mel#9 (Cordaux 2008) but could not be assembled or spanned with Sanger sequencing in the *w*Yak genome. White gene models (arrows) indicate a partial transposase gene present in the WO prophage island, truncated in *w*Yak. Pairwise differences were calculated using a sliding window (window size = 200 bp, step size = 25 bp).

The absence of *cin* loci in *w*Mel, combined with the similarity of this region between *w*Yak and distantly related *w*Pip, led us to assess how these Wolbachia acquired *cin* loci. Targeted assembly extended the scaffold *w*Yak "702380" containing the *cin* loci (Figure 5) but did not definitively place it relative to the *w*Mel genome. BLAST searches indicated that contig *w*Yak "187383" was likely flanked by *w*Yak "702380". PCR primers designed in both contigs amplified the intervening region (labeled "A" in Figure 5), confirming we have discovered the first WO prophage with two sets of *cif* loci. Subsequent Sanger sequencing revealed that this region contains an ISWpi1 element (Supplemental Information), found in many Wolbachia genomes, but not present at this location in the *w*Mel reference sequence (Cordeaux *et al*. 2008).

597 (Unpublished long-read Nanopore data shared with us by Antonio Serrato-Capuchina and
598 Jeremy Wang fully support our PCR and Sanger sequence results in the "A" region of Fig 5.)
599 IS elements encode a transposase gene that mediates their movement (Chandler and
600 Mahillon 2002). ISWpi1 elements are related to the IS5 family, which seems to be restricted to
601 Wolbachia (Duron *et al.* 2005; Cordaux *et al*. 2008). ISWpi1 elements occur in more than half of
602 the Wolbachia strains that have been evaluated, including *w*Yak and *w*Mel (Cordaux *et al.* 2008);
603 and these elements occur at variable copy number, potentially facilitating horizontal transmission
604 of ISWpi1 elements and their intervening sequence between Wolbachia variants/strains (Cordaux
605 *et al*. 2008). Following placement of contig *w*Yak "702380" relative to *w*Mel, we aligned these
606 regions and calculated pairwise differences along the chromosome using a sliding window.
607 In addition to containing the *cinA-cinB* loci, absent in *w*Mel, contig *w*Yak "702380" is on
608 average about 10% different from *w*Mel. In contrast, most of *w*Yak is less than 1% different
609 from *w*Mel (for instance, across the 650,559 bp used in our phylogenetic analyses, *w*Mel and
610 *w*Yak differ by only 0.09%). Downstream of the *cinA-cinB*$^{w\text{Yak-clade}}$ region, targeted assembly
611 enabled us to join several more contigs. The junctions were corroborated by mapping paired-end
612 reads (Langmead and Salzberg 2012) and visually inspecting the resulting bam files around
613 joined contigs for reads spanning the new junctions and for concordant read-pair mappings.
614 However, our attempts to fully bridge the gap downstream of the *cinA-cinB*$^{w\text{Yak-clade}}$ genes via
615 targeted assembly, scaffolding, and PCR were unsuccessful (see "*w*Yak unassembled gap" in
616 Figure 5). The unassembled region in *w*Yak contains an ISWpi1 element in *w*Mel, Mel #9
617 (labeled "B" in Figure 5; Cordaux 2008). Although not part of our *w*Yak assembly, homologs of
618 this ISWpi1 element appear in assemblies of this region of the WO prophage in several other A-
619 group Wolbachia (Cordaux *et al.* 2008; Bordenstein and Bordenstein 2016), specifically *w*Inc
620 from *D. incompta* (Wallau *et al.* 2016) and *w*Ri from *D. simulans* (see Figure 2). In both, we find
621 orthologs to Mel#9 ISWpi1 with corresponding flanking sequence assembled, indicating that this
622 IS element probably occurs in this unassembled region of *w*Yak, *w*San, and *w*Tei. The *w*Yak
623 sequence between these two ISWpi1 elements is highly diverged from *w*Mel in comparison to
624 the rest of the *w*Yak genome (about 10% difference versus 1%). It therefore seems plausible that
625 this region experienced a horizontal transfer event in the ancestor of *w*Yak, *w*Tei, and *w*San,
626 mediated by the flanking ISWpi1 elements.

627

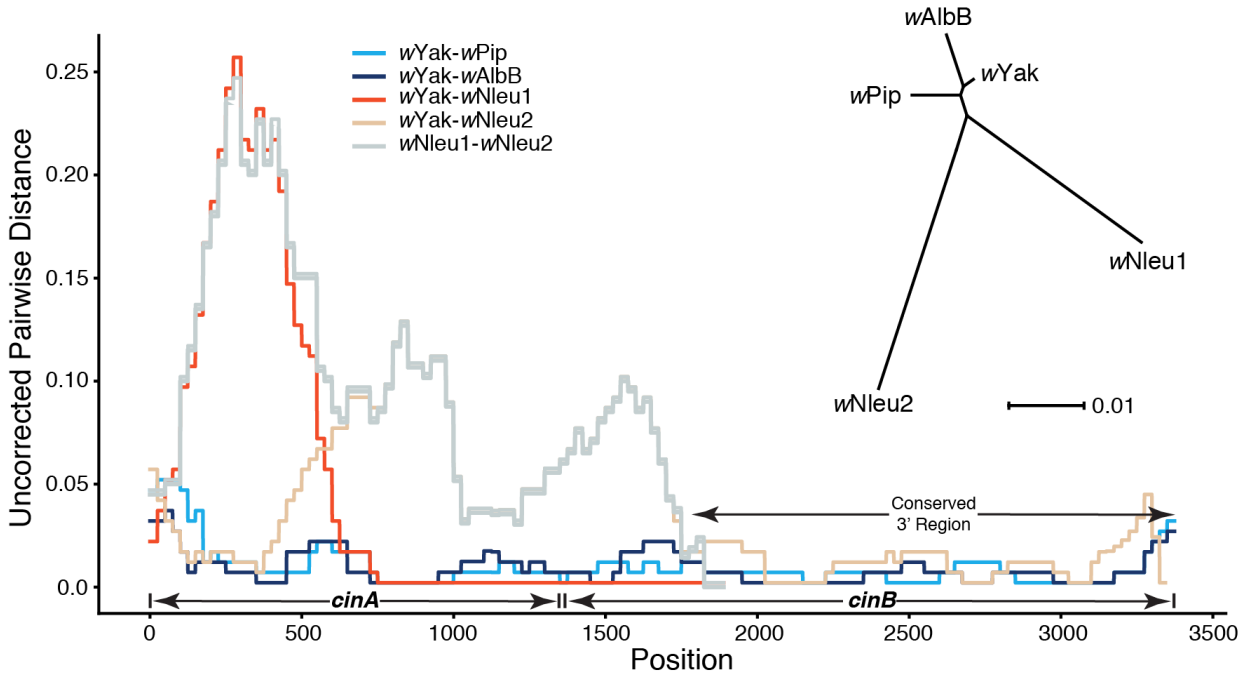

Figure 6 The observed pairwise differences between *cinA* and *cinB* copies in *w*Yak and those in *w*AlbB and *w*Pip, not including flanking regions, are shown. We also plot the pairwise differences between the two copies of these loci found in Nomada Wolbachia (denoted *w*Nleu1 and *w*Nleu2); we chose *w*Nleu to represent the Nomada-clade Wolbachia because this strain contains the longest assembled contigs for both *cinA*-*cinB* copies. Homologous *cinA*-*cinB* copies have relatively low divergence (inset unrooted tree), with the highest divergence in the 5' end of region. The most highly diverged region among any of the copies is the first ~750 bp of *w*Nleu1, with pairwise differences between *w*Nleu1 and *w*Yak reaching ~25% in some windows, with similar divergence between *w*Nleu1 and *w*Nleu2 in this same region. Pairwise differences abruptly change 750 bp from the start of the *cinA* gene with the remaining ~1,150 bp of the assembled region having a single difference from the *w*Yak sequence (only ~1,900 bp assembled into a single contig in the *w*Nleu1 copy). This pattern suggests a recent transfer/recombination event from the same unknown donor of the *w*Yak *cinA*-*cinB* copy. The unrooted tree was generated using RAxML 8.2.9 (Stamatakis 2014), and the representative *cinA*-*cinB* gene copies were obtained using BLAST to search all Wolbachia genomes in the NCBI assemblies database (https://www.ncbi.nlm.nih.gov/assembly/?term=wolbachia).

628 We conjectured that horizontal transfer occurred via the excision of the two ISWpi1
629 elements and the intervening DNA from the donor followed by homologous recombination
630 within the IS elements. To assess the plausibility of this scenario, we used BLAST to compare
631 *cinA*-*cinB*^{wPip} genes against all published *Wolbachia* genomes in the NCBI assembly database

632 (https://www.ncbi.nlm.nih.gov/assembly/?term=wolbachia). We found homologs of the *cinA-*
633 *cinB*$^{w\text{Pip}}$ in the group-A Wolbachia associated with Nomada bees (Gerth and Bleidorn 2016). An
634 unrooted tree of the ORFs for *cinA* and *cinB* (Figure 6) indicates that these genes in *w*Yak, *w*San,
635 and *w*Tei are more similar to *cinA-cinB* from group-B *w*Pip and *w*AlbB than to *cinA-cinB* found
636 in the fellow group-A Wolbachia associated with Nomada bees (*w*NFe, *w*NPa, *w*NLeu, and
637 *w*NFa), which each harbor two different *cinA-cinB* copies. This indicates that *cinA-cinB*$^{w\text{Yak-clade}}$
638 were acquired via a horizontal transfer event of these CI loci across Wolbachia groups B and A
639 that is independent of the event(s) that placed *cinA-cinB* in the Wolbachia associated with
640 Nomada bees, suggesting repeated transfers of *cin* operons.

641 The two *cinA-cinB* copies (denoted *w*NLeu1 and *w*NLeu2 in Figure 6) in Nomada
642 Wolbachia are nearly as distinct from each other as they are from the homologs in *w*Pip, *w*AlbB,
643 and *w*Yak (~7% diverged from these strains, Figure 6). However, among the four Wolbachia-
644 infected Nomada species, the orthologs are very similar with *cinA*$^{w\text{NLeu1}}$ having only 0–0.15%
645 pairwise differences among the four strains and *cinA*$^{w\text{NLeu2}}$ having 0–0.56% pairwise differences.
646 (Reconstructions for *cinB* gene copies were more complicated as the *cinB*$^{w\text{NLeu1}}$ copy fails to
647 assemble into a single contig on the 3-prime end.) This pattern suggests that *w*NLeu1 and
648 *w*NLeu2 *cin* copies were acquired by the common ancestor of the four Nomada Wolbachia
649 strains analyzed, followed by cladogenic transfer across Nomada species (Gerth and Bleidorn
650 2016). The highly fragmented assemblies of the four Nomada Wolbachia strains, with duplicate
651 copies confounding assembly, make it difficult to determine the relative positions of the *cinA* and
652 *cinB* copies and if they are likewise flanked by ISWpi1 elements.

653 To determine whether these genes were potentially moved by ISWpi1 elements into the
654 Wolbachia of the *D. yakuba* clade and Nomada, we searched each genome using both the *cinA-*
655 *cinB*$^{w\text{Yak-clade}}$ contig and the flanking ISWpi1 elements. Long repeated elements like ISWpi1
656 (916bp) break most short-read assemblies. Despite this, there is often a small fragment of the
657 element, the length of the short read, on either end of the broken contig, indicative of the repeat
658 element being responsible for terminating contig extension. We looked for the footprint of these
659 elements at the edges of the contigs on which the *cin* genes were found. We found ISWpi1
660 elements in the region flanking the both *cinA*$^{w\text{NLeu1}}$ and *cinA*$^{w\text{NLeu2}}$ copies, consistent with our
661 *w*Yak assembly in which we verified the ISWpi1 element with Sanger sequencing. These data
662 support a role for ISWpi1 in the acquisition of the *cinA-cinB* genes by the Wolbachia in the *D.*

663 *yakuba* clade and the Nomada bees. We conjecture that future work will fully confirm ISWpi1 in
664 the horizontal movement of incompatibility loci between Wolbachia.
665
666 **DISCUSSION**
667 Our results indicate introgressive and horizontal Wolbachia acquisition in the *D. yakuba* clade.
668 Evidence for horizontal Wolbachia transfer here and elsewhere (Turelli *et al.* 2018) suggests that
669 double infections must be common, even if ephemeral. Such double infections provide an
670 opportunity for ISWpi1 transposable elements to mediate horizontal transfer of incompatibility
671 loci among divergent Wolbachia. Importantly, our results highlight that incompatibility factors
672 may move independently of prophage, as evidenced by our discovery of the first prophage
673 documented to have two sets of *cif* loci. We discuss these conclusions below.
674
675 **mtDNA Introgression**
676 Our relative mitochondrial chronogram provides strong support for three mtDNA clades,
677 including a monophyletic *D. teissieri* clade that is outgroup to two sister clades: one clade
678 consisting of mtDNA from 14 *D. santomea* individuals, the other contains all *D. yakuba* mtDNA
679 plus mtDNA from 12 *D. santomea* individuals. Our results suggest less mitochondrial
680 introgression in the *D. yakuba* clade than a past report that used sequence data from only two
681 mitochondrial loci (*COII* and *ND5*, 1777 bp) and reported a clade with mtDNA sampled from
682 each species represented (Figure 3 in Bachtrog *et al*. 2006). (Using sequence data from only
683 *COII* and *ND5*, we can replicate this result, indicating data from additional loci are needed to add
684 resolution.) Our results also agree with recent work that demonstrated little nuclear introgression
685 in the *D. yakuba* clade (Turissini and Matute 2017).
686 Our results broadly agree with the results produced by Llopart *et al.* (2014) who assessed
687 mitochondrial introgression between *D. santomea* and *D. yakuba* using whole mitochondrial
688 genomes. They generated a neighbor-joining tree that produced a clade consisting of all *D.*
689 *yakuba* individuals and 10 *D. santomea* individuals, and this clade is sister to a clade with 6 *D.*
690 *santomea* individuals; one *D. santomea* haplotype is outgroup to all other haplotypes included in
691 their analysis (Figure 3 in Llopart *et al.* 2014). Nested within the mixed *D. yakuba* clade, Llopart
692 *et al.* (2014) identified a "hybrid zone clade" that includes *D. yakuba* individuals sampled from
693 São Tomé and from continental Africa and *D. santomea* individuals sampled from both within

694 and outside the Pico de São Tomé hybrid zone (HZ). The sister *D. santomea* clade also contains
695 both HZ and non-HZ individuals. Thus, their analysis and ours provide support for hybridization
696 within and outside of the HZ, leading us to question both the existence of a HZ clade and the
697 claim that these species "share the same mitochondrial genome" (Llopart *et al.* 2014). Instead,
698 both our results and theirs suggest unidirectional introgression of *D. yakuba* mitochondria into *D.*
699 *santomea*; we find 59% of *D. santomea* individuals having *D. yakuba*-like mitochondria and they
700 found 46%.
701 Llopart *et al.* (2014) used a strict molecular clock to estimate the mitochondrial MRCA of *D.*
702 *santomea* and *D. yakuba* at 10,792-17,888 years by calibrating their tree with the *D. yakuba-D.*
703 *erecta* split, estimated at 10.4 mya (Tamura *et al.* 2004). A high level of mitochondrial saturation
704 over time, with an expected value of 1.44 substitutions per synonymous site for the *D. yakuba-D.*
705 *erecta* split, could influence this estimate (Llopart *et al.* 2014). Moreover, Ho *et al.* (2005)
706 demonstrated that the mitochondrial substitution rate resembles an exponential curve, with high
707 short-term substitution rates that approach the mutation rate, then slow to a long-term rate after
708 about 1–2 million years of divergence, far younger than the inferred *D. yakuba-D. erecta* nuclear
709 and mtDNA divergence. Hence, using the slow long-term *D. yakuba-D. erecta* calibration is
710 likely to underestimate the more rapid rates of divergence experienced by *D. yakuba*-clade
711 mtDNA, inflating their short-term divergence-time estimates (Ho *et al*. 2005). If we assume that
712 Wolbachia and mtDNA were transferred by introgression, which our analyses support, our
713 estimates in Figure 3 suggest that *D. santomea* and *D. yakuba* mtDNA diverged more recently,
714 with our point estimates ranging from about 1,500 to 2,800 years.
715
716 **Wolbachia placement, divergence, and acquisition**
717 **Wolbachia placement:** Despite their similarity, *w*San, *w*Tei, and *w*Yak form monophyletic
718 clades with *w*Tei outgroup to sisters *w*San and *w*Yak, recapitulating host relationships (Figure 1;
719 Turissini and Matute 2017). *w*Mel is sister to the *D. yakuba*-clade strains in the A group (Figure
720 2), which also includes *D. simulans* strains (*w*Ri, *w*Au, and *w*Ha), *w*Ana, *w*Suz, and the Nomada
721 Wolbachia (*w*NFa, *w*NLeu, *w*NPa, and *w*NFe). Our phylogram (Figure 2A) places *w*AlbB
722 outgroup to *w*No and *w*Pip strains that diverged from A-group Wolbachia about 6–46 mya
723 (Meany *et al.* 2018).
724

725 **Wolbachia divergence:** Our chronogram analyses (Figure 3) estimate that *D. yakuba*-clade
726 Wolbachia and three *w*Mel variants diverged about 29,000 years ago, and that *w*Tei split from
727 sisters *w*San and *w*Yak about 2,500–4,500 years ago, with *w*San and *w*Yak diverging about
728 1,600-2,800 years ago. We estimate that the two most divergent *w*Mel variants from Richardson
729 *et al.* (2012) and the reference *w*Mel genome split about 4,900–7,200 years ago, indicating more
730 divergence among *w*Mel variants than among *D. yakuba*-clade Wolbachia strains. All of these
731 results depend on the calibration provided by Richardson *et al.* (2012) and the relative accuracy
732 of the underlying models of molecular evolution, which assume constant relative rates of change
733 across data partitions. For the deepest divergence in Figure 3, between *D. yakuba*-clade
734 Wolbachia and *w*Mel, we find that the estimated time depends on the variance in our prior for
735 substitution rates across branches, with a strict clock putting the divergence at about 73,000 years
736 rather than 29,000 years obtained with the most variable prior. Despite this uncertainty, the
737 quantitative differences of our Wolbachia divergence-time estimates do not alter the qualitative
738 conclusion that these Wolbachia did not co-diverge with these hosts, which split several million
739 years ago.
740
741 **Wolbachia acquisition—introgression versus horizontal:** Our divergence-time estimates for
742 the *D. yakuba*-clade Wolbachia versus their hosts preclude cladogenic acquisition. Unlike
743 introgression, horizontal (or paternal) Wolbachia acquisition should produce discordance
744 between phylogenies inferred for Wolbachia and the associated mtDNA. With the notable
745 exception of *D. santomea* line Quija 37, which has mtDNA belonging to the clade associated
746 with *D. yakuba*, but has *w*San Wolbachia, we find no evidence of discordance between the
747 estimated mtDNA and Wolbachia phylogenies. Hence our data indicate that acquisition by
748 introgression is far more common than horizontal transmission between closely related species,
749 consistent with data on acquisition of *w*Ri-like Wolbachia in the *D. melanogaster* species group
750 (Turelli *et al*. 2018). Similarly, consistent with extensive data from *D. melanogaster* (Richardson
751 *et al*. 2012) and *D. simulans* (Turelli *et al*. 2018) and smaller samples from *D. suzukii* and *D.*
752 *ananassae* (Turelli *et al.* 2018), we find only one possible example of paternal transmission or
753 horizontal transmission within *D. yakuba*-clade species.
754 We also investigated an alternative approach to distinguish between introgressive and
755 horizontal Wolbachia acquisition by estimating substitution ratios for mtDNA versus Wolbachia

(Turelli *et al.* 2018). Because we could estimate this ratio on each branch, we conjectured that this approach might have greater resolving power than our incompletely resolved mtDNA and Wolbachia phylogenies. We expect higher ratios with horizontal transmission because mtDNA would have been diverging longer than recently transferred Wolbachia. However, this approach assumes that mtDNA and Wolbachia substitution rates remain relatively constant. This is contradicted by the finding of Ho *et al.* (2005) that mtDNA substitution rates decline substantially with increasing divergence time, reaching an asymptote after around 1–2 million years. To calibrate their Wolbachia substitution rate estimates, Richardson *et al.* (2012) used an experimentally observed mitochondrial mutation rate in *D. melanogaster* (6.2×10^{-8} mutations per third-position site per generation) that extrapolates to 62% third-position divergence per million years. This extrapolation is nonsensical for a long-term substitution rate. As summarized by Ho *et al*. (2005), typical mtDNA substitution rates are 0.5–1.5% per coding site per million years. Nevertheless, using the ratio of short-term rates for mtDNA and Wolbachia, Richardson *et al.* (2012) produced an estimate of the long-term Wolbachia substitution rate that agrees with independent estimates from Nasonia wasps (Raychoudhury *et al.* 2009) and Nomada bees (Gerth and Bleidorn 2016) derived from much longer divergence times (assuming cladogenic Wolbachia acquisition) (Conner *et al*. 2017). This paradox is resolved if Wolbachia molecular evolution is not subject to the dramatic slowdown in rates seen for mtDNA substitution rates.

The apparent difference between mtDNA molecular evolution (dramatic slowdown over longer time scales, Ho *et al*. 2005) and Wolbachia molecular evolution (relative constancy, as inferred from similar rates of differentiation over very different time scales) suggests why our relative-rate test does not reject introgressive transmission of Wolbachia between *D. melanogaster* and the *D. yakuba* clade, even though it is clearly impossible. A roughly 50-fold slowdown in mtDNA substitution rates over the time scale of the divergence of *D. melanogaster* from the *D. yakuba* clade, relative to the rate of differentiation within the *D. yakuba* clade, produces comparable mtDNA-Wolbachia substitution ratios for comparisons within the *D. yakuba* clade and between *D. melanogaster* and the *D. yakuba* clade. Because of this complication and our conjecture that relative rates of mtDNA versus Wolbachia substitutions over longer periods are likely to mirror the ten-fold differences we see for mtDNA versus nuclear genes, phylogenetic discordance between mtDNA and Wolbachia is clearly a much more robust indicator of horizontal (or paternal) Wolbachia transmission. Nevertheless, additional

787 examples of cladogenic Wolbachia acquisition are needed to better understand relative rates and
788 patterns of Wolbachia, mtDNA and nuclear differentiation over different time scales.
789 Our divergence-times estimate of the *D. yakuba*-clade Wolbachia versus their hosts precludes
790 cladogenic transmission; and our phylogenetic analyses suggest that these species share very
791 similar Wolbachia because of introgression, as originally argued by Lachaise *et al.* (2000).
792 However, under either introgressive or horizontal transfer of Wolbachia, we expect the donor
793 species Wolbachia sequences would appear paraphyletic when analyzed jointly with the
794 Wolbachia from the recipient. Paraphyly allowed Turelli *et al*. (2018) to infer that *D. simulans*
795 likely obtained its Wolbachia from *D. ananassae* and *D. subpulchrella* likely obtained its
796 Wolbachia from *D. suzukii*. Paraphyly is generally expected soon after gene flow stops between
797 populations. As noted by Hudson and Coyne (2002, Fig. 1), the time scale expected to produce
798 reciprocal monophyly for mtDNA (and Wolbachia) under a neutral model of molecular evolution
799 is on the order of the effective size of the species. Our results in Figure 3 indicate that, at least for
800 our small samples, reciprocal monophyly for the Wolbachia in these three species has been
801 achieved within a few thousand years. This suggests that reciprocal monophyly has been
802 accelerated by species-specific selective sweeps within the Wolbachia or mtDNA of these
803 species. This conjecture may be testable from estimates of host fitness using transinfected versus
804 native Wolbachia.
805
806 **IS transposable elements mediate horizontal transfer of incompatibility loci between**
807 **divergent Wolbachia**
808 Wolbachia in all three *D. yakuba*-clade hosts cause both intra-and interspecific CI (Cooper *et al.*
809 2017), despite originally being characterized as non-CI causing (Zabalou *et al.* 2004; Charlat *et*
810 *al.* 2004). CI is relatively weak, and its strength can vary among *w*Tei variants and *D. teissieri*
811 backgrounds (Table 3 in Cooper *et al.* 2017). Differences in CI among Wolbachia variants has
812 also been demonstrated in interspecific backgrounds where *w*Tei caused stronger CI in a *D.*
813 *simulans* background (97.2 ± 1.3 SE percent embryo mortality) than either *w*Yak (26.5 ± 4.2 SE,
814 percent embryo mortality) or *w*San (24.0 ± 4.1 SE, percent embryo mortality) (Zabalou et al.
815 2008). Both *w*Yak and *w*San induced CI in *D. simulans* comparable in intensity to that found by
816 Cooper *et al.* (2017, Figure 3) in their original hosts. Surprisingly, Zabalou *et al*. (2008) found
817 that the strength of CI induced by *w*Tei in *D. simulans* even eclipsed that of *w*Ri (89.8 ± 4.5 SE

818 percent embryo mortality). These results are hard to reconcile with the fact that loci known to
819 underlie CI do not vary within or among *D. yakuba*-clade Wolbachia variants we examined. The
820 nearly complete CI induced by *w*Tei in *D. simulans* may depend on CI-causing factors yet to be
821 identified or differences in gene expression.

822 In each *D. yakuba*-clade Wolbachia variant included in our analyses, we find a disruption of
823 *cidB*$^{w\text{Yak-clade}}$ with an inversion from amino acids 37–103 relative to the same region in sister
824 *w*Mel. The inversion introduces multiple stop codons that could render this gene nonfunctional.
825 Fixation of loss-of-function mutations in CI-causing loci is consistent with theoretical analyses
826 showing that selection on Wolbachia within host lineages does not act to increase or maintain CI
827 (Prout 1994; Turelli 1994; Haygood and Turelli 2009); indeed, we have also recently observed a
828 single mutation that disrupts *cidB* in non-CI causing *w*Mau Wolbachia that infect *D. mauritiana*
829 on the island of Mauritius (Meany et al. 2018). In both *w*Mau and the *D. yakuba*-clade
830 Wolbachia, we find fixation of defects in the putative toxin gene. We expect that future genomic
831 analyses will produce additional examples.

832 All *D. yakuba*-clade Wolbachia genomes included in our analysis harbor *cinA-cinB* loci
833 originally discovered in the *w*Pip strain that diverged from A-group Wolbachia, including the *D.*
834 *yakuba*-clade variants, about 6–46 mya (Meany *et al*. 2018). *cin* loci are also present in B-group
835 *w*AlbB that infects *Ae. albopictus* and in A-group *w*NFe, *w*NPa, *w*NLeu, and *w*NFa Wolbachia
836 that infect Nomada bees. *cin* loci are absent from *w*Mel, but the *w*Yak contig containing these
837 loci is about 10% diverged from *w*Mel, while observed divergence between *w*Yak and *w*Mel
838 across the rest of the genome is less than 1%. The *w*Yak-clade *cin* loci share about 97%
839 similarity with the divergent B-group *w*Pip strain. *w*Yak-clade *cin* loci are more similar to *cinA-*
840 *cinB* from the B-group Wolbachia *w*Pip and *w*AlbB than to those in A-group Nomada Wolbachia
841 strains, which have two sets of *cin* loci that are as divergent from each other as they are from
842 these regions in *w*Yak and in B-group Wolbachia. These observations suggest independent
843 horizontal transfer of *cin* loci into *w*Yak and Nomada Wolbachia.

844 Our results indicate that independent of prophage movement, ISWpi1-element paralogs
845 serve to move incompatibility loci as transposon cargo via the excision of flanking ISWpi1
846 elements, followed by homologous recombination within the elements. Horizontal Wolbachia
847 acquisition is common in Drosophila (Turelli *et al.* 2018) and other species (O'Neill *et al.* 1992),
848 suggesting that double infections, which could provide the opportunity for ISWpi1 mediated

849 transfer of incompatibility loci, may be common, even if transient. (A second infection need not
850 become stably transmitted for horizontal gene transfer via ISWpi1 elements to occur.) In
851 contrast, phage particles or virions could be introduced by a vector and provide the opportunity
852 for ISWpi1 mediated transfer (Ahmed *et al.* 2015; Brown and Lloyd 2015), without the presence
853 of a double Wolbachia infection. Determining whether the insertion of these loci was derived
854 from a prophage region of the Wolbachia genome, or from a phage genome encapsulated in a
855 phage particle, remains an open question. While the ISWpi1 element in *w*Mel (Mel #9 labeled
856 "B" in Figure 5; Cordaux 2008) is not part of our *w*Yak assembly, homologs of this element are
857 present in assemblies of several other A-group Wolbachia including *w*Inc and *w*Ri (Cordaux *et*
858 *al.* 2008; Bordenstein and Bordenstein 2016). We predict this element occurs in the unassembled
859 region of our *w*Yak assembly. Footprints of ISWpi1 elements in the region flanking the *cinA*
860 genes for both copies of the gene in the Nomada Wolbachia provide further support for our
861 hypothesis. Long-read-based Wolbachia assemblies from many infected host systems will
862 elucidate the role of ISWpi1 elements in horizontal transfer of CI loci. Overall, the ecology of
863 horizontal Wolbachia transmission is crucial to understanding Wolbachia acquisition, and the
864 transfer and dynamics of CI loci are crucial to understanding Wolbachia evolution.
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879

880 **LITERATURE CITED**

881 Ahmed, M. Z., S.-J. Li, X. Xue, X.-J. Yin, S.-X. Ren *et al.*, 2015 The Intracellular Bacterium
882 Wolbachia Uses Parasitoid Wasps as Phoretic Vectors for Efficient Horizontal
883 Transmission. PLoS Pathogens 11**:** e1004672.
884 Bachtrog, D., K. Thornton, A. Clark and P. Andolfatto, 2006 Extensive introgression of
885 mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group.
886 Evolution 60**:** 292-302.
887 Bakovic V., M. Schebeck, A. Telschow, C. Stauffer, H. Schuler, 2018 Spatial spread
888 of Wolbachia in *Rhagoletis cerasi* populations. Biol. Lett. 14:20180161
889 Baldo, L., N. A. Ayoub, C. Y. Hayashi, J. A. Russell, J. K. Stahlhut *et al.*, 2008 Insight into the
890 routes of Wolbachia invasion: high levels of horizontal transfer in the spider genus
891 Agelenopsis revealed by Wolbachia strain and mitochondrial DNA diversity. Mol. Ecol.
892 17**:** 557-569.
893 Baldo, L., J. C. D. Hotopp, K. A. Jolley, S. R. Bordenstein, S. A. Biber *et al.*, 2006 Multilocus
894 sequence typing system for the endosymbiont *Wolbachia pipientis*. Appl. Environ.
895 Microbiol. 72**:** 7098-7110.
896 Bandi, C., T. J. Anderson, C. Genchi and M. L. Blaxter, 1998 Phylogeny of Wolbachia in filarial
897 nematodes. Proc. Biol. Sci. 265**:** 2407-2413.
898 Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin *et al.*, 2012 SPAdes: a new
899 genome assembly algorithm and its applications to single-cell sequencing. J. Comput.
900 Biol. 19**:** 455-477.
901 Beckmann, J. F., M. Bonneau, H. Chen, M. Hochstrasser, D. Poinsot *et al.*, The Toxin &
902 Antidote Model of Cytoplasmic Incompatibility: Genetics and Evolutionary Implications.
903 Trends Genet. In press.
904 Beckmann, J. F., and A. M. Fallon, 2013 Detection of the Wolbachia protein WPIP0282 in
905 mosquito spermathecae: implications for cytoplasmic incompatibility. Insect Biochem.
906 Mol. Biol. 43**:** 867-878.
907 Beckmann, J. F., J. A. Ronau and M. Hochstrasser, 2017 A Wolbachia deubiquitylating enzyme
908 induces cytoplasmic incompatibility. Nat. Microbiol. 2: 17007.
909 Bordenstein, S. R., and S. R. Bordenstein, 2016 Eukaryotic association module in phage WO
910 genomes from Wolbachia. Nat. Commun. 7**:** 13155.
911 Bouchon, D., T. Rigaud and P. Juchault, 1998 Evidence for widespread Wolbachia infection in
912 isopod crustaceans: molecular identification and host feminization. Proc. Biol. Sci. 265:
913 1081-1090.
914 Braig, H. R., W. Zhou, S. L. Dobson and S. L. O'Neill, 1998 Cloning and characterization of a
915 gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia*
916 *pipientis*. J. Bacteriol. 180: 2373-2378.
917 Brand, C. L., S. B. Kingan, L. Wu and D. Garrigan, 2013 A Selective Sweep across Species
918 Boundaries in Drosophila. Mol. Biol. Evol. 30: 2177-2186.
919 Brown, A. N., and V. K. Lloyd, 2015 Evidence for horizontal transfer of Wolbachia by a
920 Drosophila mite. Exp. Appl. Acarol. 66: 301-311.
921 Brownlie, J. C., B. N. Cass, M. Riegler, J. J. Witsenburg, I. Iturbe-Ormaetxe *et al.*, 2009
922 Evidence for Metabolic Provisioning by a Common Invertebrate Endosymbiont,
923 *Wolbachia pipientis*, during Periods of Nutritional Stress. PLoS Pathogens 5**:** e1000368.
924 Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+:
925 architecture and applications. BMC Bioinformatics 10**:** 421.

Carson, H. L., K. Y. Kaneshiro and F. C. Val, 1989 Natural hybridization between the sympatric Hawaiian species *Drosophila silvestris* and *Drosophila heteroneura*. Evolution 43**:** 190-203.

Chandler, M., and J. Mahillon, 2002 Insertion Sequences Revisited in *Mobile DNA II*. American Society of Microbiology.

Charlat, S., J. W. O. Ballard and H. Merçot, 2004 What maintains noncytoplasmic incompatibility inducing Wolbachia in their hosts: a case study from a natural *Drosophila yakuba* population. Journal of Evolutionary Biology 17**:** 322-330.

Chiu, J. C., X. Jiang, L. Zhao, C. A. Hamm, J. M. Cridland *et al.*, 2013 Genome of *Drosophila suzukii*, the spotted wing Drosophila. G3 3**:** 2257-2271.

Clark, A. G., M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver *et al.*, 2007 Evolution of genes and genomes on the Drosophila phylogeny. Nature 450**:** 203-218.

Comeault, A. A., A. Venkat and D. R. Matute, 2016 Correlated evolution of male and female reproductive traits drive a cascading effect of reinforcement in *Drosophila yakuba*. Proc. Biol. Sci. 283: 20160730.

Conner, W. R., M. L. Blaxter, G. Anfora, L. Ometto, O. Rota-Stabelli *et al.*, 2017 Genome comparisons indicate recent transfer of *w*Ri-like Wolbachia between sister species *Drosophila suzukii* and *D. subpulchrella*. Ecol. Evol.**:** 1-14.

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.

Cooper, B. S., A. Sedghifar, W. T. Nash, A. A. Comeault and D. R. Matute, 2018 A Maladaptive Combination of Traits Contributes to the Maintenance of a Drosophila Hybrid Zone. Curr. Biol. 28**:** 2940-2947.e2946.

Cordaux, R., 2008 ISWpi1 from *Wolbachia pipientis* defines a novel group of insertion sequences within the IS5 family. Gene 409**:** 20-27.

Coyne, J. A., S. Y. Kim, A. S. Chang, D. Lachaise and S. Elwyn, 2002 Sexual isolation between two sibling species with overlapping ranges: *Drosophila santomea* and *Drosophila yakuba*. Evolution 56**:** 2424-2434.

Coyne, J. A., and H. A. Orr, 1989 Patterns of speciation in Drosophila. Evolution 43**:** 362-381.

Coyne, J. A., and H. A. Orr, 1997 "Patterns of Speciation in Drosophila" Revisited. Evolution 51**:** 295-303.

Duron, O., J. Lagnel, M. Raymond, K. Bourtzis, P. Fort *et al.*, 2005 Transposable element polymorphism of Wolbachia in the mosquito *Culex pipiens*: evidence of genetic diversity, superinfection and recombination. Mol. Ecol. 14**:** 1561-1573.

Ellegaard, K. M., L. Klasson, K. Näslund, K. Bourtzis and S. G. E. Andersson, 2013 Comparative Genomics of Wolbachia and the Bacterial Species Concept. PLoS Genetics 9**:** e1003381.

Garrigan, D., S. B. Kingan, A. J. Geneva, P. Andolfatto, A. G. Clark *et al.*, 2012 Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. Genome Res. 22**:** 1499-1511.

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

Gill, A. C., A. C. Darby and B. L. Makepeace, 2014 Iron necessity: the secret of Wolbachia's success? PLoS Negl. Trop. Dis. 8**:** e3224-e3224.

Giordano, R., S. L. O'Neill and H. M. Robertson, 1995 Wolbachia infections and the expression of cytoplasmic incompatibility in *Drosophila sechellia* and *D. mauritiana*. Genetics 140**:** 1307-1317.

Gloor, G. B., C. R. Preston, D. M. Johnsonschlitz, N. A. Nassif, R. W. Phillis *et al.*, 1993 Type-1 repressors of P-element mobility. Genetics 135**:** 81-95.

Hamm, C. A., D. J. Begun, A. Vo, C. C. R. Smith, P. Saelao *et al.*, 2014 Wolbachia do not live by reproductive manipulation alone: infection polymorphism in *Drosophila suzukii* and *D. subpulchrella*. Mol. Ecol. 23**:** 4871-4885.

Havird, J. C., and D. B. Sloan, 2016 The Roles of Mutation, Selection, and Expression in Determining Relative Rates of Evolution in Mitochondrial versus Nuclear Genomes. Mol. Biol. Evol. 33**:** 3042-3053.

Haygood, R., and M. Turelli, 2009 Evolution of incompatibility-inducing microbes in subdivided host populations. Evolution 63**:** 432-447.

Hedges, L. M., J. C. Brownlie, S. L. O'Neill and K. N. Johnson, 2008 Wolbachia and virus protection in insects. Science 322: 702.

Hilgenboecker, K., P. Hammerstein, P. Schlattmann, A. Telschow and J. H. Werren, 2008 How many species are infected with Wolbachia? A statistical analysis of current data. FEMS Microbiol. Lett. 281**:** 215-220.

Ho, S. Y., M. J. Phillips, A. Cooper and A. J. Drummond, 2005 Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol. Biol. Evol. 22**:** 1561-1568.

Hoffmann, A. A., 1988 Partial cytoplasmic incompatibility between two Australian populations of *Drosophila melanogaster*. Entomol. Exp. Appl. 48**:** 61-67.

Hoffmann, A. A., D. Clancy and J. Duncan, 1996 Naturally-occurring Wolbachia infection in *Drosophila simulans* that does not cause cytoplasmic incompatibility. Heredity 76**:** 1-8.

Hoffmann, A. A., D. J. Clancy and E. Merton, 1994 Cytoplasmic incompatibility in Australian populations of *Drosophila melanogaster*. Genetics 136: 993-999.

Hoffmann, A. A., B. L. Montgomery, J. Popovici, I. Iturbe-Ormaetxe, P. H. Johnson *et al.*, 2011 Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. Nature 476**:** 454-U107.

Hoffmann, A. A., and M. Turelli, 1988 Unidirectional incompatibility in *Drosophila simulans*: inheritance, geographic variation and fitness effects. Genetics 119**:** 435-444.

Hoffmann, A. A., and M. Turelli, 1997 Cytoplasmic incompatibility in insects, pp. 42–80 in Influential Passengers: Inherited Microorganisms and Arthropod Reproduction, edited by S. L. O'Neill, A. A. Hoffmann, and J. H. Werren. Oxford University Press, New York.

Hoffmann, A. A., M. Turelli and G. M. Simmons, 1986 Unidirectional incompatibility between populations of *Drosophila simulans*. Evolution 40**:** 692-701.

Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot *et al.*, 2016 RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language. Syst. Biol. 65: 726-736.

Hoskins, R. A., J. W. Carlson, K. H. Wan, S. Park, I. Mendez *et al.*, 2015 The Release 6 reference sequence of the *Drosophila melanogaster* genome. Genome Res. 25**:** 445-458.

Hudson, R. R. and J. A. Coyne, 2002 Mathematical consequences of the genealogical species concept. Evolution 56:1557–1565.

1016 Huigens, M. E., R. P. de Almeida, P. A. Boons, R. F. Luck and R. Stouthamer, 2004 Natural
1017 interspecific and intraspecific horizontal transfer of parthenogenesis-inducing Wolbachia
1018 in Trichogramma wasps. Proc. Biol. Sci. 271**:** 509-515.
1019 Huigens, M. E., R. F. Luck, R. H. Klaassen, M. F. Maas, M. J. Timmermans *et al.*, 2000
1020 Infectious parthenogenesis. Nature 405**:** 178-179.
1021 Hunter, S. S., R. T. Lyon, B. A. J. Sarver, K. Hardwick, L. J. Forney *et al.*, 2015 Assembly by
1022 Reduced Complexity (ARC): a hybrid approach for targeted assembly of homologous
1023 sequences. bioRxiv**:** 014662.
1024 Hurst, G. D. D., and F. M. Jiggins, 2000 Male-killing bacteria in insects: mechanisms, incidence,
1025 and implications. Emerg. Infect. Dis. 6**:** 329-336.
1026 Jackman, S. D., B. P. Vandervalk, H. Mohamadi, J. Chu, S. Yeo *et al.*, 2017 ABySS 2.0:
1027 resource-efficient assembly of large genomes using a Bloom filter. Genome Res. 27**:** 768-
1028 777.
1029 Jaenike, J., K. A. Dyer, C. Cornish and M. S. Minhas, 2006 Asymmetrical reinforcement and
1030 Wolbachia infection in Drosophila. PLoS Biol. 4**:** 1852-1862.
1031 Joshi, N. A., and J. N. Fass, 2011 Sickle: A sliding-window, adaptive, quality-based trimming
1032 tool for FastQ files, pp.
1033 Kamimura, Y., and H. Mitsumoto, 2012 Lock-and-key structural isolation between sibling
1034 Drosophila species. Entomol. Sci. 15**:** 197-201.
1035 Katoh, K., and D. M. Standley, 2013 MAFFT Multiple Sequence Alignment Software Version 7:
1036 Improvements in Performance and Usability. Mol. Biol. Evol. 30**:** 772-780.
1037 Kern, A. D., D. A. Barbash, J. Chang Mell, A. Jensen and D. Hupalo, 2015 Highly Constrained
1038 Intergenic Drosophila Ultraconserved Elements Are Candidate ncRNAs. Genome Biol.
1039 Evol. 7**:** 689-698.
1040 Klasson, L., T. Walker, M. Sebaihia, M. J. Sanders, M. A. Quail *et al.*, 2008 Genome evolution
1041 of Wolbachia strain *w*Pip from the *Culex pipiens* group. Mol. Biol. Evol. 25**:** 1877-1887.
1042 Klasson, L., J. Westberg, P. Sapountzis, K. Näslund, Y. Lutnaes *et al.*, 2009 The mosaic genome
1043 structure of the Wolbachia *w*Ri strain infecting *Drosophila simulans*. PNAS 106**:** 5725-
1044 5730.
1045 Kondo, R., Y. Satta, E. T. Matsuura, H. Ishiwa, N. Takahata *et al.*, 1990 Incomplete maternal
1046 transmission of mitochondrial DNA in Drosophila. Genetics 126**:** 657-663.
1047 Kriesner, P., W. R. Conner, A. R. Weeks, M. Turelli and A. A. Hoffmann, 2016 Persistence of a
1048 Wolbachia infection frequency cline in *Drosophila melanogaster* and the possible role of
1049 reproductive dormancy. Evolution 70**:** 979-997.
1050 Kriesner, P., A. A. Hoffmann, S. F. Lee, M. Turelli and A. R. Weeks, 2013 Rapid Sequential
1051 Spread of Two Wolbachia Variants in *Drosophila simulans*. PLoS Path. 9.
1052 Kulathinal, R. J., L. S. Stevison and M. A. F. Noor, 2009 The Genomics of Speciation in
1053 Drosophila: Diversity, Divergence, and Introgression Estimated Using Low-Coverage
1054 Genome Sequencing. PLoS Genet. 5**:** e1000550.
1055 Lachaise, D., M. Harry, M. Solignac, F. Lemeunier, V. Benassi *et al.*, 2000 Evolutionary
1056 novelties in islands: *Drosophila santomea*, a new *melanogaster* sister species from São
1057 Tomé. Proc. Sci. Biol. 267**:** 1487-1495.
1058 Langley, C. H., and W. M. Fitch, 1974 An examination of the constancy of the rate of molecular
1059 evolution. J. Mol. Evol. 3**:** 161-177.
1060 Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nature
1061 Methods 9**:** 357-359.

1062 Laven, H., 1951 Crossing Experiments with Culex Strains. Evolution 5**:** 370-375.
1063 Laven, H., 1967 Eradication of *Culex pipiens fatigans* through cytoplasmic incompatibility.
1064 Nature 216**:** 383-384.
1065 LePage, D. P., J. A. Metcalf, S. R. Bordenstein, J. On, J. I. Perlmutter *et al.*, 2017 Prophage WO
1066 genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility. Nature
1067 543**:** 243-247.
1068 Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping
1069 and population genetical parameter estimation from sequencing data. Bioinformatics 27**:**
1070 2987-2993.
1071 Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler
1072 transform. Bioinformatics 25**:** 1754-1760.
1073 Li, S.-J., M. Z. Ahmed, N. Lv, P.-Q. Shi, X.-M. Wang *et al.*, 2017 Plantmediated horizontal
1074 transmission of Wolbachia between whiteflies. ISME Journal 11**:** 1019-1028.
1075 Lindsey, A. R. I., D. W. Rice, S. R. Bordenstein, A. W. Brooks, S. R. Bordenstein *et al.*, 2018
1076 Evolutionary Genetics of Cytoplasmic Incompatibility Genes *cifA* and *cifB* in Prophage
1077 WO of *Wolbachia*. Genome Biol. Evol. 10**:** 434-451.
1078 Llopart, A., D. Herrig, E. Brud and Z. Stecklein, 2014 Sequential adaptive introgression of the
1079 mitochondrial genome in *Drosophila yakuba* and *Drosophila santomea*. Mol. Ecol. 23**:**
1080 1124-1136.
1081 Lohse, K., M. Clarke, M. G. Ritchie and W. J. Etges, 2015 Genome-wide tests for introgression
1082 between cactophilic Drosophila implicate a role of inversions during speciation.
1083 Evolution 69**:** 1178-1190.
1084 Martinez, J., B. Longdon, S. Bauer, Y.-S. Chan, W. J. Miller *et al.*, 2014 Symbionts Commonly
1085 Provide Broad Spectrum Resistance to Viruses in Insects: A Comparative Analysis of
1086 Wolbachia Strains. PLoS Path. 10**:** e1004369.
1087 Matute, D. R., 2010 Reinforcement of gametic isolation in Drosophila. PLoS Biol. 8**:** e1000341.
1088 Matute, D. R., and J. F. Ayroles, 2014 Hybridization occurs between *Drosophila simulans* and
1089 *D. sechellia* in the Seychelles archipelago. J. Evol. Biol. 27**:** 1057-1068.
1090 Matute, D. R., and J. A. Coyne, 2010 Intrinsic reproductive isolation between two sister species
1091 of Drosophila. Evolution 64**:** 903-920.
1092 Mavingui, P., C. Valiente Moro, V. Tran-Van, F. Wisniewski-Dye, V. Raquin *et al.*, 2012
1093 Whole-genome sequence of Wolbachia strain *w*AlbB, an endosymbiont of tiger mosquito
1094 vector *Aedes albopictus*. J. Bacteriol. 194**:** 1840.
1095 McMeniman, C. J., R. V. Lane, B. N. Cass, A. W. C. Fong, M. Sidhu *et al.*, 2009 Stable
1096 introduction of a life-shortening Wolbachia infection into the mosquito *Aedes aegypti*.
1097 Science 323**:** 141-144.
1098 Meany, M. K., W. R. Conner, S. V. Richter, J. A. Bailey, M. Turelli *et al.*, 2018 Loss of
1099 cytoplasmic incompatibility and minimal fecundity effects explain relatively low
1100 Wolbachia frequencies in *Drosophila mauritiana*. bioRxiv**:** 461574.
1101 Moriyama, M., N. Nikoh, T. Hosokawa and T. Fukatsu, 2015 Riboflavin Provisioning Underlies
1102 Wolbachia's Fitness Contribution to Its Insect Host. mBio 6. e01732-15.
1103 Nosil, P., 2013 Degree of sympatry affects reinforcement in Drosophila. Evolution 67**:** 868-872.
1104 Nurk, S., A. Bankevich, D. Antipov, A. Gurevich, A. Korobeynikov *et al.*, 2013 Assembling
1105 Genomes and Mini-metagenomes from Highly Chimeric Reads, pp. 158-170 in *Research*
1106 *in Computational Molecular Biology*, edited by M. Deng, R. Jiang, F. Sun and X. Zhang.
1107 Springer Berlin Heidelberg, Berlin, Heidelberg.

1108 O'Connor, L., C. Plichart, A. C. Sang, C. L. Brelsfoard, H. C. Bossin *et al.*, 2012 Open release of
1109 male mosquitoes infected with a Wolbachia biopesticide: field performance and infection
1110 containment. PLoS Negl. Trop. Dis. 6**:** e1797.
1111 O'Neill, S. L., R. Giordano, A. M. Colbert, T. L. Karr and H. M. Robertson, 1992 16S rRNA
1112 phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic
1113 incompatibility in insects. PNAS 89**:** 2699-2702.
1114 O'Neill, S. L., A. A. Hoffman and J. H. Werren, 1998 *Influential Passengers: Inherited*
1115 *Microorganisms and Arthropod Reproduction*. Oxford University Press.
1116 Prout, T., 1994 Some Evolutionary Possibilities for a Microbe that Causes Incompatibility in its
1117 Host. Evolution 48**:** 909-911.
1118 Raychoudhury, R., L. Baldo, D. C. S. G. Oliveira and J. H. Werren, 2009 Modes of acquisition
1119 of Wolbachia: horizontal transfer, hybrid introgression, and codivergence in the Nasonia
1120 species complex. Evolution 63**:** 165-183.
1121 Richardson, M. F., L. A. Weinert, J. J. Welch, R. S. Linheiro, M. M. Magwire *et al.*, 2012
1122 Population genomics of the Wolbachia endosymbiont in *Drosophila melanogaster*. PLoS
1123 Genet. 8**:** e1003129.
1124 Ritchie S. A., 2018 Wolbachia and the near cessation of dengue outbreaks in Northern Australia
1125 despite continued dengue importations via travellers. J. Travel Med. 25:tay084.
1126 Rousset, F., D. Bouchon, B. Pintureau, P. Juchault and M. Solignac, 1992 Wolbachia
1127 endosymbionts responsible for various alterations of sexuality in arthropods. Proc. Biol.
1128 Sci. 250: 91-98.
1129 Rousset, F., and M. Solignac, 1995 Evolution of single and double Wolbachia symbioses during
1130 speciation in the *Drosophila simulans* complex. PNAS 92**:** 6389-6393.
1131 Sánchez, L., and P. Santamaria, 1997 Reproductive isolation and morphogenetic evolution in
1132 Drosophila analyzed by breakage of ethological barriers. Genetics 147**:** 231-242.
1133 Schmidt, T. L., N. H. Barton, G. Rašić, A. P. Turley, B. L. Montgomery *et al.*, 2017 Local
1134 introduction and heterogeneous spatial spread of dengue-suppressing Wolbachia through
1135 an urban population of *Aedes aegypti*. PLoS Biol. 15**:** e2001894.
1136 Schrider, D. R., J. Ayroles, D. R. Matute and A. D. Kern, 2018 Supervised machine learning
1137 reveals introgressed loci in the genomes of *Drosophila simulans* and *D. sechellia*. PLoS
1138 Genet. 14**:** e1007341.
1139 Schuler, H., K. Köppler, S. Daxböck-Horvath, B. Rasool, S. Krumböck *et al.*, 2016 The
1140 hitchhiker's guide to Europe: the infection dynamics of an ongoing Wolbachia invasion
1141 and mitochondrial selective sweep in *Rhagoletis cerasi*. Mol. Ecol. 25**:** 1595-1609.
1142 Seemann, T., 2014 Prokka: rapid prokaryotic genome annotation. Bioinformatics 30**:** 2068-2069.
1143 Shoemaker, D. D., V. Katju and J. Jaenike, 1999 Wolbachia and the evolution of reproductive
1144 isolation between *Drosophilla recens* and *Drosophila subquinaria*. Evolution 53**:** 1157-
1145 1164.
1146 Shropshire, J. D., J. On, E. M. Layton, H. Zhou and S. R. Bordenstein, 2018 One prophage WO
1147 gene rescues cytoplasmic incompatibility in *Drosophila melanogaster*. PNAS 115: 4987-
1148 4991.
1149 Simao, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva and E. M. Zdobnov, 2015
1150 BUSCO: assessing genome assembly and annotation completeness with single-copy
1151 orthologs. Bioinformatics 31**:** 3210-3212.
1152 Stamatakis, A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of
1153 large phylogenies. Bioinformatics (Oxford, England) 30**:** 1312-1313.

1154 Sutton, E. R., S. R. Harris, J. Parkhill and S. P. Sinkins, 2014 Comparative genome analysis of
1155 Wolbachia strain *w*Au. BMC Genomics 15**:** 928.
1156 Teixeira, L., A. Ferreira and M. Ashburner, 2008 The bacterial symbiont Wolbachia induces
1157 resistance to RNA viral infections in *Drosophila melanogaster*. PLoS Biol. 6**:** e1000002.
1158 Turelli, M., 1994 Evolution of incompatibility-inducing microbes and their hosts. Evolution 48.
1159 Turelli, M., B. S. Cooper, K. M. Richardson, P. S. Ginsberg, B. Peckenpaugh *et al.*, 2018 Rapid
1160 Global Spread of *w*Ri-like Wolbachia across Multiple Drosophila. Curr. Biol. 28**:** 963-
1161 971.e968.
1162 Turelli, M., and A. A. Hoffmann, 1991 Rapid spread of an inherited incompatibility factor in
1163 California Drosophila. Nature 353**:** 440-442.
1164 Turelli, M., and A. A. Hoffmann, 1995 Cytoplasmic incompatibility in *Drosophila simulans*:
1165 dynamics and parameter estimates from natural populations. Genetics 140**:** 1319-1338.
1166 Turelli, M., J. R. Lipkowitz and Y. Brandvain, 2014 On the Coyne and Orr-igin of species:
1167 effects of intrinsic postzygotic isolation, ecological differentiation, X chromosome size,
1168 and sympatry on Drosophila speciation. Evolution 68**:** 1176-1187.
1169 Turissini, D. A., and D. R. Matute, 2017 Fine scale mapping of genomic introgressions within
1170 the Drosophila yakuba clade. PLoS Genet. 13**:** e1006971.
1171 Vavre, F., L. Mouton and B. A. Pannebakker, 2009 Chapter 12 *Drosophila*–Parasitoid
1172 Communities as Model Systems for Host–*Wolbachia* Interactions, pp. 299-331 in
1173 *Advances in Parasitology*. Academic Press, San Diego, California.
1174 Wallau, G. L., M. T. da Rosa, F. C. De Re and E. L. Loreto, 2016 Wolbachia from *Drosophila*
1175 *incompta*: just a hitchhiker shared by Drosophila in the New and Old World? Insect Mol.
1176 Biol. 25**:** 487-499.
1177 Weeks, A. R., M. Turelli, W. R. Harcombe, K. T. Reynolds and A. A. Hoffmann, 2007 From
1178 parasite to mutualist: rapid evolution of Wolbachia in natural populations of Drosophila.
1179 PLoS Biol. 5**:** e114.
1180 Weinert, L. A., E. V. Araujo-Jnr, M. Z. Ahmed and J. J. Welch, 2015 The incidence of bacterial
1181 endosymbionts in terrestrial arthropods. Proc. Biol. Sci. 282**:** 20150249.
1182 Werren, J. H., and D. M. Windsor, 2000 Wolbachia infection frequencies in insects: evidence of
1183 a global equilibrium? Proc. Biol. Sci. 267**:** 1277-1285.
1184 Werren, J. H., W. Zhang and L. R. Guo, 1995 Evolution and phylogeny of Wolbachia:
1185 reproductive parasites of arthropods. Proc. Biol. Sci. 261**:** 55-63.
1186 Wu, M., L. V. Sun, J. Vamathevan, M. Riegler, R. Deboy *et al.*, 2004 Phylogenomics of the
1187 reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by
1188 mobile genetic elements. PLoS Biol. 2**:** e69.
1189 Yen, J. H., and A. R. Barr, 1971 New hypothesis of the cause of cytoplasmic incompatibility in
1190 *Culex pipiens L*. Nature 232**:** 657-658.
1191 Yukilevich, R., 2012 Asymmetrical patterns of speciation uniquely support reinforcement in
1192 Drosophila. Evolution 66**:** 1430-1446.
1193 Zabalou, S., A. Apostolaki, S. Pattas, Z. Veneti, C. Paraskevopoulos *et al.*, 2008 Multiple rescue
1194 factors within a Wolbachia strain. Genetics 178**:** 2145-2160.
1195 Zabalou, S., S. Charlat, A. Nirgianaki, D. Lachaise, H. Merçot *et al.*, 2004 Natural Wolbachia
1196 infections in the *Drosophila yakuba* species complex do not induce cytoplasmic
1197 incompatibility but fully rescue the *w*Ri modification. Genetics 167**:** 827-834.
1198

1199

1200 **SUPPLEMENTAL INFORMATION**

1201

1202 **SUPPLEMENTAL MATERIALS AND METHODS**

1203 **PCR amplification and Sanger Sequencing of "A" and "B" regions in Figure 5.**

1204

1205 DNA was extracted from each isofemale (yakuba_S09_L24, yakuba_S01_L19, and
1206 yakuba_S01_L31) line using a standard 'squish' buffer protocol (Gloor *et al*. 1993). A standard
1207 PCR assay (Simpliamp ThermoCycler, Applied Biosystems, Singapore) was performed using the
1208 GoTaq™ DNA Polymerase (Promega™ , Wisconsin, USA) master mix. PCR program began
1209 with 3 minutes at 94°C, followed by 34 cycles of 30 seconds at 94°C, 30 seconds at 55°C, and 1
1210 minute and 15 seconds at 72°C. The profile finished with 8 minutes at 72°C. The amplified
1211 PCR products were visualized on a 1% agarose gel that included a 1kb ladder.

1212 For each fly line, each of the three yak_ISWpi1 primer sets were assayed (Table S1). CinA
1213 primer set consistently returned a 100-150 bp product, 2CinA returned a ~1000 bp product, and
1214 692607_694583 primer set returned no products. Touchdown PCR with an annealing gradient
1215 from 70°C to 55°C was performed with the same results to the standard PCR run from above.

1216 For each fly line, both Mel_ISWPi1 primer sets were assayed (Table S1). All assays
1217 returned no products. Touchdown PCR with an annealing gradient from 70°C to 55°C was
1218 performed and also returned no products. Annealing temp was lowered from 55°C to 50°C and
1219 the PCR was repeated, and again no products were returned.

1220 The ~1000 bp PCR product from yakuba_S09_L24 was cleaned using the QiaQuick PCR
1221 purification kit (Qiagen, Germany). The cleaned PCR product was sent to Eurofins (Kentucky)
1222 for Sanger sequencing.

1223
1224
1225
1226
1227
1228
1229
1230
1231

1232 **SUPPLEMENTAL TABLES**

1233 **Table S1.** Primers used in this study.

Locus/Target region	Primer name	Forward primer	Reverse primer	Reference
wsp: Wolbachia surface protein	*wsp*	5'-TGGTCCAATAAGTGATGAAGAAAC-3'	5'-AAAAATTAAACGCTACTCCA-3'	Braig *et. al.* 1998
2L Control: rDNA	*2L*	5'-TGCAGCTATGGTCGTTGACA-3'	5'-ACGAGACAATAATATGTGGTGCTG-3'	Meany *et al.* 2018
YAK_***IS****Wpi1*, "A" in Figure 5§	2CinA	5'-TTTTTGTTCTCAGAGGTTGTCC -3'	5'-ATCCAACCATTTCCAGCAAC -3'	Designed here
YAK_***IS****Wpi1*, "A" in Figure 5‡	CinA	5'-TTTTAGCATGAGAAAGCAAGTCC-3'	5'-TTTTCCAACTTCTGTGCAAATG-3'	Designed here
YAK_***IS****Wpi1*, "A" in Figure 5*	692607_2_694583	5'-CCTTTATTGGCAATGCTGCT -3'	5'-CCTAAATTTTATTACGGCACTTTCA -3'	Designed here
MEL_***IS****Wpi1*, "B" in Figure 5*	692607_2_433806	5'-CCTCAGGTTCGAGTGACAATTT -3'	5'-TAAGTTGTCTTCAATATGCCGC -3'	Designed here
MEL_***IS****Wpi1*, "B" in Figure 5*	433808_2_694583	5'-GCTTACCAAGGCATGAATCC -3'	5'-TGCTTATTTTTCTTGTGGTCCA -3'	Designed here

1234 §Produced a ~1000 bp product
1235 ‡Produced a ~100-150 bp product
1236 *Failed to amplify
1237
1238

Table S2. The scaffold count, N50, and total assembly size of each Wolbachia pseudoreference.

Genome	Scaffold count	N50	Total assembly size
*w*Yak	118	17888	1.28 million bases
*w*San	300	7231	1.18 million bases
*w*Tei	122	16157	1.30 million bases

1239

1240 **Table S3.** Near-universal, single-copy proteobacteria genes used to assess assembly
1241 completeness. We used BUSCO v. 3.0.0 to identify complete and fragmented proteobacteria
1242 genes (out of 221) found in our draft Wolbachia genomes. For comparison, we provide results
1243 for reference-quality Wolbachia genomes.

Genome	Complete	Fragment	Not found
Complete reference genomes			
*w*Ri	180	2	39
*w*Mel	180	2	39
*w*Au	181	2	38
*w*Ha	179	3	39
*w*No	181	4	36
Draft genome assemblies			
*w*Yak	181	2	38
*w*San	181	2	38
*w*Tei	180	2	39

1244
1245
1246
1247
1248
1249
1250

Table S4. Rate multipliers for the Wolbachia relaxed clock chronograms. All numbers are estimated from the Γ(2,2) trees. They are the total estimated substitutions per site for that partition across the relevant clade (See Methods).

	1st position (95% credible interval)	**2nd position** (95% credible interval)	**3rd position** (95% credible interval)
D. yakuba-Wolbachia	1.9×10^{-5}, (9.6×10^{-6}, 3.0×10^{-5})	1.2×10^{-5}, (5.2×10^{-6}, 1.9×10^{-5})	1.7×10^{-5}, (8.5×10^{-6}, 2.7×10^{-5})
*w*Mel	4.4×10^{-5}, (2.2×10^{-5}, 7.1×10^{-5})	5.7×10^{-5}, (2.8×10^{-5}, 9.2×10^{-5})	3.4×10^{-5}, (1.5×10^{-5}, 5.6×10^{-5})
D. yakuba- clade and *w*Mel Wolbachia	1.7×10^{-4}, (1.1×10^{-4}, 2.4×10^{-4})	1.4×10^{-4}, (8.7×10^{-5}, 1.9×10^{-4})	2.0×10^{-4}, (1.2×10^{-4}, 2.7×10^{-4})

1251