Introgressive and horizontal acquisition of Wolbachia by hybridizing *Drosophila yakuba*-clade hosts and horizontal transfer of phage 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 phage loci transfer in the *D. yakuba*-clade

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

15 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 species ((*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 very similar 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 mitochondria 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 more recent common ancestors than the bulk of the host nuclear genomes, precluding cladogenic Wolbachia acquisition. A single discordance between mitochondrial and Wolbachia phylogenies indicates rare horizontal or paternal transmission across close relatives. General concordance of Wolbachia and mitochondrial phylogenies suggest that horizontal transmission is rare, but statistical tests are complicated by varying relative rates of molecular divergence. WO prophage loci associated with CI in *w*Mel are disrupted in the *D. yakuba*-clade Wolbachia, but these A-group Wolbachia contain a second set of CI loci, horizontally transferred from distantly related B-group Wolbachia via transposable elements.

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 loci cause CI (Beckmann and Fallon 2013; Beckmann *et al*. 2017; LePage *et al*. 2017). 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

53 nematodes (Bandi *et al*. 1998); and there are also examples in at least two insect clades, Nasonia
54 wasps (Raychoudhury *et al.* 2009) and Nomada bees (Gerth and Bleidorn 2016). In contrast,
55 infections can be relatively young and acquired through introgressive or horizontal transfer
56 (Raychoudhury *et al.* 2009; Schuler *et al*. 2016; Conner *et al.* 2017; Turelli *et al.* 2018).
57 Comparisons of host nuclear and mitochondrial genomes with the associated Wolbachia
58 genomes enable discrimination among these three modes (Raychoudhury *et al.* 2009; Turelli *et*
59 *al.* 2018). Concordant nuclear, mitochondrial, and Wolbachia cladograms—including consistent
60 divergence-time estimates for all three genomes—support cladogenic acquisition and co-
61 divergence of host and Wolbachia lineages. Concordant Wolbachia and mtDNA phylogenies and
62 consistent Wolbachia and mitochondrial divergence-time estimates that are more recent than
63 nuclear divergence support introgressive acquisition. In this case, mitochondrial and Wolbachia
64 relationships may or may not recapitulate the host phylogeny. Finally, if Wolbachia diverged
65 more recently than either nuclear or mitochondrial genomes, horizontal transfer (or paternal
66 transmission; Hoffmann and Turelli 1988; Turelli and Hoffmann 1995) is indicated. This is often
67 associated with discordance between host and Wolbachia phylogenies (O'Neill *et al.* 1992;
68 Rousset and Solignac 1995; Turelli *et al.* 2018).
69 Introgressive Wolbachia acquisition may be common in Drosophila. About half of all
70 closely related Drosophila species have overlapping geographical ranges and show pervasive
71 evidence of reinforcement (Coyne and Orr 1989, 1997; Yukelevich 2012; Nosil 2013), indicating
72 that hybridization must be common (Turelli *et al.* 2014). Several instances of sporadic
73 contemporary hybridization and interspecific gene flow have been documented in the genus
74 (Carson *et al.* 1989; Shoemaker *et al.* 1999; Jaenike *et al.* 2006; Kulathinal *et al.* 2009; Garrigan
75 *et al.* 2012; Brand *et al.* 2013; Matute and Ayroles 2014; Lohse *et al.* 2015); but only two stable
76 hybrid zones have been well described, and both involve *D. yakuba*-clade species. In west
77 Africa, *D. yakuba* hybridizes with endemic *D. santomea* on the island of São Tomé, and with *D.*
78 *teissieri* on the island of Bioko (Lachaise *et al.* 2000; Llopart *et al.* 2005; Comeault *et al.* 2016;
79 Cooper *et al.* 2018). The ranges of *D. yakuba* and *D. teissieri* overlap throughout continental
80 Africa, but contemporary hybridization has not been observed outside of Bioko (Cooper *et al.*
81 2018). Genomic analyses support both mitochondrial and nuclear introgression in the *D. yakuba*
82 clade (Lachaise *et al.* 2000; Bachtrog *et al.* 2006; Llopart *et al.* 2014; Turissini and Matute 2017;
83 Cooper *et al.* 2018), and Wolbachia infecting all three species (*w*San, *w*Tei, and *w*Yak) are at

intermediate frequencies and identical with respect to commonly used typing loci (Lachaise *et al*. 2000; Charlat *et al*. 2004; Cooper *et al.* 2017).

Wolbachia target the germ line to ensure vertical transmission following horizontal transfer (Frydman *et al.* 2006; Pietri *et al.* 2016). Some closely related bacteria–Rickettsia (Azad *et al.* 1998; Baldridge *et al.* 2004), Holosporaceae (Kaltz and Koella 2003), and Neorickettsia (Gibson *et al.* 2005)–display both vertical and horizontal transmission (reviewed in Perlman *et al.* 2006). But most related bacteria (*e.g*., Anaplasma, Ehrlichia, and Bartonella) are not transmitted transovarially (Munderloh and Kurtti 1995; Long *et al.* 2003; Perlman *et al.* 2006). Horizontal Wolbachia transmission has been repeatedly demonstrated since its initial discovery by O'Neill *et al*. (1992) (*e.g*., Baldo *et al.* 2008; Schuler *et al.* 2016), and it may be common in some systems (*e.g*., Huigens *et al.* 2000; Huigens *et al.* 2004; Ahmed *et al*. 2015; Li *et al.* 2017). Phylogenomic data indicate recent horizontal Wolbachia transmission, on the order of 5,000–27,000 years, among relatively distantly related Drosophila species of the *D. melanogaster* species group (Turelli *et al.* 2018); but there is little evidence for non-maternal transmission within Drosophila species (Richardson *et al.* 2012; Turelli *et al.* 2018). Horizontal Wolbachia transfer could occur via a vector (Vavre *et al*. 2009; Ahmed *et al*. 2015) and/or via shared food sources during development (Huigens *et al.* 2000; Li *et al*. 2017). Rare paternal Wolbachia transmission has been documented in *D. simulans* (Hoffmann and Turelli 1988; Turelli and Hoffmann 1995). However, the most common mode of Wolbachia acquisition by Drosophila species (and most other hosts) remains unknown, and all three modes seem plausible in the *D. yakuba* clade (Lachaise *et al*. 2000; Bachtrog *et al*. 2006; Cooper *et al*. 2017).

Distinguishing between acquisition via introgression, non-sexual horizontal, or paternal transmission requires estimating phylogenies and relative divergence times of mtDNA and Wolbachia (Raychoudhury *et al.* 2009; Conner *et al*. 2017; Turelli *et al*. 2018). To convert sequence divergence to divergence times requires understanding relative rates of divergence for mtDNA, nuclear genomes and Wolbachia. If each genome followed a constant-rate molecular clock, taxa with cladogenic Wolbachia transmission, such as Nasonia wasps (Raychoudhury *et al.* 2009) and Nomada bees (Gerth and Bleidorn 2016), would provide reference rates for calibration. However, like Langley and Fitch (1974), Turelli *et al*. (2018) found significantly varying relative rates of Wolbachia and mtDNA divergence. This variation, which we document more extensively, confounds attempts to understand Wolbachia acquisition, as discussed below.

115 Here, we use host and Wolbachia genomes from the *D. yakuba* clade to demonstrate
116 introgressive and horizontal Wolbachia acquisition, but general concordance of mtDNA and
117 Wolbachia phylogenies indicates that horizontal acquisition is rare within this clade. However,
118 tests involving divergence-time estimates are complicated by varying relative rates and patterns
119 of Wolbachia, mtDNA, and nuclear sequence divergence, as illustrated by data from more
120 distantly related Drosophila (Clark *et al.* 2007). Finally, we demonstrate that phage loci
121 underlying CI in closely related *w*Mel are disrupted in all *D. yakuba*-clade Wolbachia (LePage *et*
122 *al.* 2017); however, these Wolbachia contain another set of CI loci, absent in *w*Mel, but very
123 similar to those found in *w*Pip (from *Culex pipiens*) (Beckmann *et al.* 2017), a strain that
124 diverged 6–46 million years ago (mya) from *w*Yak and *w*Mel (Werren *et al*. 1995; Meany *et al.*
125 2018). Several analyses implicate Wolbachia-specific insertion sequence (IS) transposable
126 elements, specifically ISWpi1, in the horizontal transfer of these CI loci and surrounding phage
127 regions between distantly related Wolbachia.
128

129 # METHODS

130

131 **Genomic data**

132 The *D. yakuba*-clade isofemale lines included in our study were sampled over several years in
133 west Africa (Comeault *et al.* 2016; Turissini and Matute 2017; Cooper *et al.* 2017). Each line
134 within each species used in our analyses exhibits little nuclear introgression (< 1%) (Turissini
135 and Matute 2017), and no hybrids were included. Reads from *D. yakuba* (*N* = 56), *D. santomea*
136 (*N* = 11), and *D. teissieri* (*N* = 13) isofemale lines were obtained from the data archives of
137 Turissini and Matute (2017) and aligned to the *D. yakuba* nuclear and mitochondrial reference
138 genomes (Clark *et al*. 2007) with bwa 0.7.12 (Li and Durbin 2009), requiring alignment quality
139 scores of at least 50. Because many of the read archives were single end, all alignments were
140 completed using single-end mode for consistency.
141

142 **mtDNA:** Consensus mitochondrial sequences for each of the 80 lines were extracted with
143 samtools v. 1.3.1 and bcftools v 1.3.1 (Li 2011). Sequences for the 13 protein-coding genes were
144 extracted, based on their positions in the *D. yakuba* reference. We also extracted the 13 protein-
145 coding genes from additional unique *D. yakuba* (*N* = 28) and *D. santomea* (*N* = 15)

146 mitochondrial genomes (Llopart *et al.* 2014), and from the *D. melanogaster* reference (Hoskins
147 *et al.* 2015). Genes were aligned using MAFFT v. 7 and concatenated (Katoh and Standley
148 2013). Lines identical across all 13 genes were represented by one sequence for the phylogenetic
149 analyses.
150
151 **Wolbachia:** Reads from each of the 80 Turissini and Matute (2017) lines were aligned using
152 bwa 0.7.12 (Li and Durban 2009) to the *D. yakuba* reference genome (Clark *et al.* 2007)
153 combined with the *w*Mel reference genome (Wu *et al.* 2004). We calculated the average depth of
154 coverage across the *w*Mel genome. We considered lines with < 1× coverage uninfected and lines
155 with >10× coverage infected. No lines had between 1× and 10× coverage. To test our genomic
156 analyses of infection status, we used a polymerase chain reaction (PCR) assay on a subset of
157 lines. We extracted DNA using a "squish" buffer protocol (Gloor *et al*. 1993), and infection
158 status was determined using primers for the Wolbachia-specific *wsp* gene (Braig *et al*. 1998;
159 Baldo *et al.* 2006). We amplified a Drosophila-specific region of chromosome *2L* as a positive
160 control (Kern *et al.* 2015) (primers are listed in Supplemental Material, Table S1). For each run,
161 we also included a known Wolbachia-positive line and a water blank as controls.
162 To produce a draft Wolbachia "pseudoreference" genome for each species, we first
163 determined the isofemale line from each species that produced the greatest average coverage
164 depth over *w*Mel. We trimmed the reads with Sickle v. 1.33 (Joshi and Fass 2011) and assembled
165 with ABySS v. 2.0.2 (Jackman *et al.* 2017). *K* values of 51, 61…91 were tried. Scaffolds with
166 best nucleotide BLAST matches to known Wolbachia sequences with E-values less than 10^{-10}
167 were extracted as the draft Wolbachia assembly. For each species, the assembly with the highest
168 N50 and fewest scaffolds was kept as our Wolbachia pseudoreference genome for that host
169 species, denoted *w*Yak, *w*San, and *w*Tei (Table S2). To assess the quality of these three draft
170 assemblies, we used BUSCO v. 3.0.0 to search for homologs of the near-universal, single-copy
171 genes in the BUSCO proteobacteria database (Simão *et al.* 2015). For comparison, we followed
172 Conner *et al*. (2017) and performed the same search on the reference genomes for *w*Mel (Wu *et*
173 *al.* 2004), *w*Ri (Klasson *et al.* 2009), *w*Au (Sutton *et al.* 2014), *w*Ha and *w*No (Ellegaard *et al.*
174 2013) (Table S3).
175 Using these draft Wolbachia pseudoreference genomes, all other samples were aligned to
176 the *D. yakuba* reference (nuclear and mitochondrial) plus the species-specific Wolbachia draft

177 assembly with bwa 0.7.12 (Li and Durbin 2009). Consensus sequences were extracted with
178 samtools v. 1.3.1 and bcftools v. 1.3.1 (Li 2011).
179 To test whether the choice of Wolbachia pseudoreference influenced the Wolbachia draft
180 sequence obtained for each infected line, we arbitrarily selected three infected lines of each host
181 species and aligned reads from each of those lines independently to our *w*Yak, *w*San, and *w*Tei
182 pseudoreferences. This resulted in 27 alignments (9 per host species). Among the three
183 Wolbachia consensus sequences generated for each line, we assessed variation across all genes
184 used for downstream analyses. Irrespective of the pseudoreference used, we obtained identical
185 Wolbachia sequences. This indicates the robustness of our approach to generating population
186 samples of draft Wolbachia genomes, without having a high-quality Wolbachia reference
187 genome from any of the host species. Because Llopart *et al.* (2014) released only assembled
188 mitochondrial genomes, we could not assess the Wolbachia infection status of their lines.
189
190 **Loci for phylogenetic and comparative-rate analyses**
191 **Wolbachia genes:** For each phylogenetic analysis of Wolbachia data, the draft genomes were
192 annotated with Prokka v. 1.11 (Seemann 2014), which identifies homologs to known bacterial
193 genes. To avoid pseudogenes and paralogs, we used only genes present in a single copy and with
194 identical lengths in all of the sequences analyzed. Genes were identified as single copy if they
195 uniquely matched a bacterial reference gene identified by Prokka v. 1.11. By requiring all
196 homologs to have identical length in all of our draft Wolbachia genomes, we removed all loci
197 with indels across any of our sequences.
198 Given that many loci accumulate indels over time, the number of loci included in our
199 phylogenetic analyses depended on the number of strains included. We first estimated
200 phylograms and chronograms for the "A group" Wolbachia from: the *D. yakuba*-clade, *w*Mel
201 from *D. melanogaster* (Wu *et al*. 2004), *w*Inc that infects *Drosophila incompta* (Wallau *et al.*
202 2016), Wolbachia that infect Nomada bees (*w*NFe, *w*NPa, *w*NLeu, and *w*NFa; Gerth and
203 Bleidorn 2016), and Wolbachia that infect *D. simulans* (*w*Ri, *w*Au and *w*Ha; Klasson *et al*. 2009;
204 Sutton *et al*. 2014; Ellegaard *et al*. 2013). We also analyzed distantly related "B group"
205 Wolbachia: *w*No from *D. simulans* (Ellegaard *et al*. 2013), *w*Pip_Pel from *Culex pipiens*
206 (Klasson *et al*. 2008), and *w*AlbB from *Aedes albopictus* (Mavingui *et al.* 2012). To increase our
207 phylogenetic resolution within the *D. yakuba* Wolbachia clade (by increasing the number of

208 loci), we estimated a phylogram that included only *D. yakuba*-clade Wolbachia and *w*Mel. Our
209 phylogram with both A- and B-group Wolbachia included 146 genes (containing 115,686 bp),
210 our chronogram with only *D. yakuba*-clade Wolbachia included 678 genes (containing 695,118
211 bp), and our phylogram that included only *D. yakuba*-clade Wolbachia and *w*Mel included 643
212 genes (containing 644,586 bp),
213 To illustrate how divergence-time estimates vary with changing patterns of molecular
214 divergence, we estimated two additional Wolbachia chronograms: one that considered only the
215 *w*Mel reference plus the two most-diverged *w*Mel lines included in Richardson *et al.* (2012), and
216 one that included these three *w*Mel variants and our *D. yakuba*-clade Wolbachia. Our
217 chronogram with only *w*Mel Wolbachia included 692 genes (containing 709,599 bp), and our
218 chronogram that included *w*Mel plus *D. yakuba*-clade Wolbachia included 621 genes (containing
219 624,438 bp).
220
221 **mtDNA and nuclear genes from diverse Drosophila:** To assess variation in the relative rates
222 of divergence of mitochondrial and nuclear genes, we analyzed the canonical 12 Drosophila
223 genomes (Clark *et al.* 2007) plus *D. suzukii* (Chiu *et al*. 2013). We excluded *D. sechellia* and *D.*
224 *persimilis* which show evidence of introgression with *D. simulans* and *D. pseudoobscura*,
225 respectively (Kulathinal *et al.* 2009; Schrider *et al.* 2018; Brand *et al.* 2013). Coding sequences
226 for 20 nuclear genes used in the analyses of Turelli *et al.* (2018) (*aconitase, aldolase, bicoid,*
227 *ebony, enolase, esc, g6pdh, glyp, glys, ninaE, pepck, pgi, pgm, pic, ptc, tpi, transaldolase, white,*
228 *wingless*, and *yellow*) were obtained from FlyBase for each species. The genes were aligned with
229 MAFFT v. 7 (Katoh and Standley 2013). Coding sequences for the 13 protein-coding
230 mitochondrial genes in the inbred reference strains were also obtained from FlyBase for each
231 species and were aligned with MAFFT v. 7 and concatenated.
232
233 **Phylogenetic analyses**
234 All of our analyses used RevBayes v. 1.0.9 (Hohna *et al.* 2016), following the procedures of
235 Turelli *et al*. (2018). For completeness, we summarize those methods below. For additional
236 details on the priors and their justifications, consult Turelli *et al*. (2018). Four independent runs
237 were performed for each analysis, and in all cases runs agreed with one another. Nodes with
238 posterior probability less than 0.95 were collapsed into polytomies.

Wolbachia phylograms: We estimated a phylogram for A- and B-group Wolbachia and for only *D. yakuba*-clade and *w*Mel Wolbachia using the same methodology as Turelli *et al.* (2018). We used a GTR + Γ model with four rate categories, partitioning by codon position. Each partition had an independent rate multiplier with prior Γ(1,1) (i.e., Exp(1)), as well as stationary frequencies and exchangeability rates drawn from flat, symmetrical Dirichlet distributions (i.e., Dirichlet(1,1,1...). The model used a uniform prior over all possible topologies. Branch lengths were drawn from a flat, symmetrical Dirichlet distribution, thus they summed to 1. Since the expected number of substitutions along a branch equals the branch length times the rate multiplier, the expected number of substitutions across the entire tree for a partition is equal to the partition's rate multiplier.

Wolbachia chronograms: We first created a relaxed-clock relative chronogram with the root age fixed to 1 using the GTR + Γ model, partitioned by codon position, using the same birth-death prior as Turelli *et al.* (2018). Each partition had an independent rate multiplier with prior Gamma(1,1), as well as stationary frequencies and exchangeability rates drawn from flat, symmetrical Dirichlet distributions. The branch-rate prior for each branch was Γ(2,2), normalized to a mean of 1 across all branches (Table S4). We also tried a strict-clock tree and a relaxed-clock tree with branch-rate prior Γ(7,7), which produced no significant differences. We used the scaled distribution $\Gamma(7,7) \times 6.87 \times 10^{-9}$ to model substitutions per third-position site per year. This transforms the relative chronogram into an absolute chronogram. This scaled distribution was chosen to replicate the upper and lower credible intervals of the posterior distribution estimated by Richardson *et al*. (2012), assuming 10 Drosophila generations per year, normalized by their median substitution-rate estimate. Branch lengths in absolute time were calculated as the relative branch length times the third position rate multiplier divided by the substitutions per third position site per year estimate above.

We illustrate how divergence-time estimates depend on changing patterns of Wolbachia molecular evolution observed over different time scales, specifically the relative rates of third-site substitutions versus first- and second-site substitutions. We compare divergence times estimated for variants within *melanogaster*-group host species, with a time scale of only hundreds or thousands of years, with the same divergence times estimated when more distantly related Wolbachia, with divergence times over tens of thousands of years, are included in the

270 analyses. We present three separate analyses: one using only *D. yakuba*-clade Wolbachia
271 variants, a second using only *w*Mel variants analyzed by Richardson *et al.* (2012), and a third
272 simultaneously analyzing both sets of data.
273
274 ***Drosophila* phylogeny for 11 references species:** We estimated a phylogram from our 20
275 nuclear loci using the GTR + Γ model, with partitioning by gene and codon position. The model
276 was identical to the Wolbachia phylogram model above except for the partitioning (there are too
277 few Wolbachia substitutions to justify partitioning by gene).
278
279 **mtDNA phylogeny for the 11 reference species**: We estimated a phylogram from the 13
280 protein-coding mitochondrial loci using the GTR + Γ model, with partitioning only by codon
281 position. The model was identical to the Wolbachia phylogram model above.
282
283 **mtDNA chronogram for the *D. yakuba* clade:** We estimated a relative chronogram for the *D.*
284 *yakuba*-clade mitochondria from the 13 protein-coding loci using the GTR + Γ model,
285 partitioning only by codon position. To test the sensitivity of our results to priors, we ran a strict
286 clock, a relaxed clock with a Γ(7,7) branch-rate prior, and a relaxed clock with a Γ(2,2) branch-
287 rate prior, corresponding to increasing levels of substitution-rate variation across branches. The
288 model was identical to the Wolbachia chronogram model above, except that we did not
289 transform it into an absolute chronogram.
290
291 **Ratios of divergence rates for mtDNA versus nuclear loci:** To quantify ratios of mtDNA to
292 nuclear substitution rates, we estimated relative substitution rates for host nuclear genes versus
293 mtDNA using the GTR + Γ model. The (unrooted) topology was fixed to the consensus topology
294 for the nuclear and mitochondrial data. The data from 20 nuclear loci were partitioned by locus
295 and codon position, the mtDNA data were partitioned only by codon position. All partitions
296 shared the same topology, but the nuclear partitions were allowed to have branch lengths
297 different from the mtDNA partitions. The sum of the branch lengths for each partition was scaled
298 to 1. Assuming concurrent nuclear-mtDNA divergence (because we used only species showing
299 no evidence of introgression), we imposed the same absolute ages for all nodes of the nuclear
300 and mtDNA chronograms. For each nuclear locus, the third-position rate-ratio for the mtDNA

301 versus nuclear genomes was calculated as: (mitochondrial branch length × mitochondrial 3rd
302 position rate multiplier)/(nuclear branch length × nuclear 3rd position rate multiplier). We
303 summarized the relative rates of mtDNA versus nuclear substitutions along each branch using
304 the arithmetic average of the 20 ratios obtained from the individual nuclear loci.
305
306 **Introgressive versus horizontal Wolbachia transfer–concordance of phylograms:** Horizontal
307 transfer of Wolbachia was initially detected by discordance between Wolbachia and host
308 phylogenies (O'Neill *et al*. 1992). When considering samples within species and closely related
309 species, a comparison of mtDNA and Wolbachia phylogenies provides a natural test for
310 horizontal transmission. We first look for significant discrepancies between strongly supported
311 nodes in the mtDNA versus Wolbachia phylogenies. To test for less obvious differences, we
312 follow Richardson *et al*. (2012) and compute Bayes Factors to assess the support for models that
313 assume that mtDNA versus Wolbachia follow the same topology versus distinct phylogenies. For
314 these calculations, we omit san-Quija37 due to its clear discordance.
315 We calculated the marginal likelihood in RevBayes for two models; one with a shared
316 mtDNA and Wolbachia phylogeny, another with independent topologies. The mtDNA and
317 Wolbachia were each partitioned by codon position, for a total of six partitions. In the shared
318 phylogeny model, all six share the same topology; in the independent model, the mtDNA and
319 Wolbachia have separate topologies. All priors were the same as the Wolbachia phylogram
320 model. Lines that were identical across the mtDNA and Wolbachia were collapsed into one
321 sample. We ran two independent replicates of each model with 50 stepping stones per run. The
322 Bayes factor is computed as the difference between the marginal likelihoods of each model.
323
324 **Introgressive versus horizontal Wolbachia transfer–relative rates:** Within species, the
325 phylogenies of mtDNA and Wolbachia are fully resolved. Thus we consider an alternative
326 approach for distinguishing between introgressive versus horizontal transfer that depends on
327 estimating divergence times for Wolbachia genomes and host mtDNA. This is complicated by
328 variation in the relative rates of mtDNA versus Wolbachia divergence and by systematic changes
329 over time in the relative rates of substitutions at the three codon positions for Wolbachia.
330 Following the procedures in Turelli *et al.* (2018), we estimated the mtDNA-to-Wolbachia third-
331 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 versus Wolbachia substitution-rate ratio for the interspecific branches separately, regardless of whether the shared rate-ratio model was favored.

Wolbachia CI-relevant loci

The discovery of weak CI in the *D. yakuba clade* (Cooper *et al*. 2017) motivates detailed comparative analysis of Wolbachia loci recently associated with causing (Beckmann and Fallon 2013; Beckmann *et al*. 2017; LePage *et al*. 2017; Lindsey *et al*. 2018) and rescuing (Shropshire *et al*. 2018) CI. Beckmann and Fallon (2013) associated wPip_0282 and wPip_0283 proteins in *w*Pip-infected *Culex pipiens* with Wolbachia-modified sperm. These proteins are produced by WO prophage genes and induce toxicity and cognate rescue when expressed/co-expressed in *Saccharomyces cerevisiae* and *D. melanogaster* (Beckmann *et al.* 2017). Homologs of these genes in *w*Mel (WD0631 and WD0632) recapitulate CI when transgenically expressed in *D. melanogaster* (LePage *et al.* 2017). A distantly related WO-prophage-associated pair, present in *w*Pip, *w*Pip_0294 (proposed antidote) and *w*Pip_0295 (proposed toxin), causes similar toxicity/cognate rescue in *S. cerevisiae*. The *w*Pip_0282-*w*Pip_0283 and the wPip_0294-wPip_0295 pairs have been named *cidA*-*cidB*$^{w\mathrm{Pip}}$ and *cinA*-*cinB*$^{w\mathrm{Pip}}$, respectively (Beckmann *et al.* 2017 and Beckmann *et al.* 2019). WD0631-WD632 have been referred to as *cifA*-*cifB* (LePage *et al*. 2017; Lindsey *et al.* 2018) and as *cidA*-*cidB*$^{w\mathrm{Mel}}$ (Beckmann *et al.* 2019). We use *cidA*-*cidB*$^{w\mathrm{Mel}}$ in this paper to avoid unnecessary confusion when comparing loci between host systems and to distinguish *cid* (CI-inducing deubiquitylating enzyme) from *cin* (CI-inducing nuclease) pairs (Beckmann *et al.* 2019). We use *cif* (CI factor) when speaking generally about CI loci.

362 We used protein BLASTN to search for all *cif* homologs in each of our *D. yakuba*-clade
363 Wolbachia assemblies, querying *cidA-cidB* variants and the paralogous pair *cinA-cinB*$^{w\mathrm{Ri}}$
364 (wRi_006720 and wRi_006710) found in *w*Ri and some *w*Ri-like Wolbachia (Beckmann and
365 Fallon 2013; Turelli *et al*. 2018), and the *cinA-cinB*$^{w\mathrm{Pip}}$ pair (Beckmann and Fallon 2013;
366 Beckmann *et al*. 2017; LePage *et al*. 2017; Lindsey *et al*. 2018). We identified both *cid* and *cin*
367 operon homologs, but we found no precise matches to the *cinA-cinB*$^{w\mathrm{Ri}}$ pair in any of our
368 genomes. For all of the samples, we extracted consensus sequences from our assemblies and
369 alignments for the *cidA-cidB*$^{w\mathrm{Yak\text{-}clade}}$ and *cinA-cinB*$^{w\mathrm{Yak\text{-}clade}}$ gene pairs. The genes were aligned
370 with MAFFT v. 7 (Katoh and Standley 2013). We examined variation in these regions relative to
371 *cidA-cidB*$^{w\mathrm{Mel}}$ and *cinA-cinB*$^{w\mathrm{Pip}}$ respectively.
372 Because we unexpectedly identified homologs of *cinA-cinB* in all *w*Yak-clade Wolbachia,
373 we took additional measures to understand the origin and placement of these loci in the genomes.
374 The *cinA-cinB*$^{w\mathrm{Yak\text{-}clade}}$ open reading frames are located on a ~11,500 bp scaffold in the
375 fragmented *w*Yak assembly (*w*Yak scaffold "702380" from ABySS output—we use quotes to
376 indicate names assigned by ABySS). The first ~4,000 bp of this scaffold contain *cinA* and *cinB*
377 genes that have ~97% identity with those in *w*Pip. We performed a BLAST search using all
378 contigs in the *w*Yak assembly as queries against the *w*Mel genome (Camacho *et al*. 2009). The
379 *cinA-cinB*$^{w\mathrm{Yak\text{-}clade}}$ scaffold was placed on the *w*Mel genome (~617,000–623,000) adjacent to and
380 downstream of the *w*Yak contig containing *cidA-cidB*$^{w\mathrm{Yak\text{-}clade}}$ loci (*w*Yak contig "187383").
381 However, only ~7,000 bp of the 11,500 bp of the scaffold containing *cinA-cinB*$^{w\mathrm{Yak\text{-}clade}}$ align to
382 this region of *w*Mel. The ~4,000 bp sequence that contains the ORFs for *cinA-cinB*$^{w\mathrm{Yak\text{-}clade}}$ has
383 no significant hit against the *w*Mel genome.
384 To verify the placement of the scaffold containing *cinA-cinB*$^{w\mathrm{Yak\text{-}clade}}$, we first performed
385 targeted assembly using the *cinA-cinB*$^{w\mathrm{Yak\text{-}clade}}$ contig and the adjacent contigs as references
386 (Hunter *et al*. 2015). Iterative targeted assembly often extends assembled scaffolds (Langmead
387 and Salzberg 2012). Reads are mapped to the target of interest, and the mapped reads are then
388 assembled using SPADES (Bankevich *et al*. 2012; Nurk *et al*. 2013). The newly assembled
389 scaffolds serve as the target in a subsequent round of mapping. The procedure is repeated until
390 no new reads are recruited to the assembled scaffold from the previous round. The extended
391 scaffolds enabled us to merge flanking *w*Yak contigs downstream of the *cinA-cinB*$^{w\mathrm{Yak\text{-}clade}}$
392 scaffold but failed to connect the fragment with the *cidA-cidB*$^{w\mathrm{Yak\text{-}clade}}$ contig. We used PCR to

393 amplify the intervening region (Sanger *et al.* 1977) (Supplemental Information, Table S1), and
394 Sanger sequencing to evaluate this product.
395 Subsequent mapping of paired-end reads to the merged scaffolds confirmed the correct
396 order and orientation of the contigs containing the *cidA-cidB*$^{w\text{Yak-clade}}$ and *cinA-cinB*$^{w\text{Yak-clade}}$
397 genes. Observed, uncorrected, pairwise distances between focal regions in *w*Yak, *w*Mel, and
398 *w*Pip and between focal regions in *w*Yak, *w*Pip, *w*AlbB, and *w*Nleu were calculated using a
399 sliding window (window size = 200 bp, step size = 25 bp).
400

401 # RESULTS

402

403 **Mitochondrial introgression**

404 Previous analyses suggest that *D. yakuba* and *D. santomea* carry very similar mtDNA due to
405 ongoing hybridization and introgression (Lachaise *et al*. 2000; Bachtrog *et al.* 2006; Llopart *et*
406 *al.* 2014). However, our mtDNA relative chronogram (Figure 1), based on mitochondrial whole-
407 proteome data sampled from throughout the ranges of all three species, supports three mtDNA
408 clades that largely agree with the nuclear topology—*D. teissieri* mtDNA sequences are outgroup
409 to sister *D. yakuba* and *D. santomea* mtDNA (Figure 1). Consistent with introgression, 12 out of
410 26 *D. santomea* isofemale lines have *D. yakuba*-like mtDNA (indicated by blue branches and
411 letters in Figure 1). Yet, our samples showed no evidence of *D. santomea* mtDNA introgressed
412 into *D. yakuba* and no mtDNA introgression involving *D. teissieri*. The 12 *D. santomea*
413 isofemales with introgressed *D. yakuba*-like mtDNA were sampled from both within (N = 8) and
414 outside (N = 4) the well-described Pico de São Tomé hybrid zone (Llopart et al. 2014; Turissini
415 and Matute 2017). The *D. santomea* hybrid-zone samples are indicated by "HZ" in Figure 1 in
416 blue, and they are found throughout the clade that includes all *D. yakuba* mtDNA. This suggests
417 that hybridization occurs in other areas of the island or that introgressed *D. santomea* genotypes
418 migrate outside of the hybrid zone.
419 Previous analysis of nuclear genomes (Turissini and Matute 2017) also found more
420 introgression from *D. yakuba* into *D. santomea* when populations of *D. yakuba* from near the
421 Gulf of Guinea, Cameroon, and Kenya were included; however, excluding Cameroon and Kenya
422 indicated similar amounts of introgression in each direction. Matings between *D.*
423 *santomea* females and *D. yakuba* males are rare relative to the reciprocal cross (Coyne *et al*.

2002, Matute 2010). When these matings do occur, F1 females produce fewer progeny than do F1 females produced by *D. yakuba* females and *D. santomea* males (Matute and Coyne 2010). F1 hybrids produced by *D. santomea* females also have a shortened lifespan (Matute and Coyne 2010), due to copulatory wounds inflicted by *D. yakuba* males during mating (Kamimura 2012). These observations are consistent with our finding of preferential introgression of *D. yakuba* mitochondria into *D. santomea* backgrounds.

Wolbachia frequencies and draft genomes

As expected, we found Wolbachia in all three *D. yakuba*-clade species sampled by Turissini and Matute (2017). For *D. yakuba,* 21 of 56 lines were infected, yielding an infection frequency of $p = 0.36$, with 95% binomial confidence interval (0.25, 0.51). For *D. santomea*, 10 of 11 were infected, $p = 0.91$ (0.59, 1.0); and for *D. teissieri*, 11 of 13 were infected, $p = 0.85$ (0.55, 0.98). Additional frequency estimates are reported in Cooper *et al*. (2017), which found that Wolbachia frequencies vary through time and space in west Africa. All PCR tests for Wolbachia matched our coverage-based genomic analyses of infection status.

The average read depth over the *w*Mel reference sequence was 1,940 for our *w*Yak pseudoreference genome (yak-CY17C), 40 for *w*San (san-Quija630.39), and 489 for *w*Tei (teis-cascade_4_2). These pseudoreference genomes were included in the phylogram that includes A- and B-group Wolbachia (Figure 2), and they cluster with the Wolbachia from their respective hosts (Figure 3), as expected if introgression is rare. In general, the *w*San data are lower quality due to the relatively small size of the *D. santomea* libraries (Turissini and Matute 2017). The scaffold count, N50, and total assembly sizes are reported in Supplementary Table S2.

Wolbachia phylograms, chronograms and relative rates

Our phylogram that includes A- and B-group Wolbachia places the three weak-CI-causing *D. yakuba*-clade Wolbachia together and sister to *w*Mel in the A group (Figure 2A). In contrast, *D. simulans* carries diverse Wolbachia that do (*w*Ri, *w*Ha, *w*No) and do not (*w*Au) cause CI, spanning Wolbachia groups A (*w*Ri, *w*Ha, *w*Au) and B (*w*No). All *w*San, *w*Yak, and *w*Tei Wolbachia variants included in our analysis are identical at the MLST loci often used to type Wolbachia variants (Cooper *et al.* 2017; Baldo *et al.* 2006). Due to the relatively small number of genes that meet our inclusion criteria (146 genes across 115,686 bp), the resulting phylogram

455

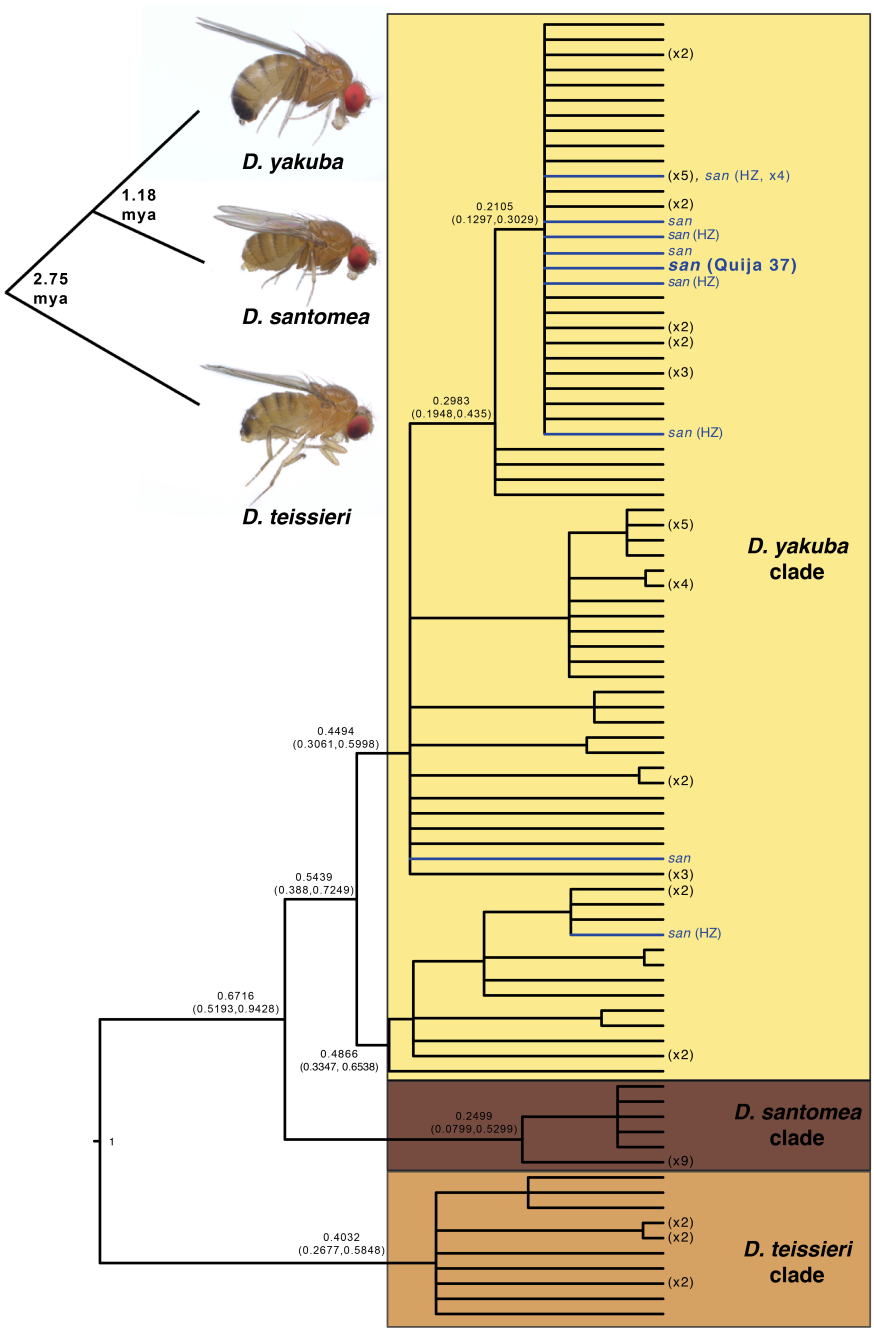

Figure 1 Bayesian relative chronogram for the *D. yakuba* clade mtDNA. The root age was fixed to 1. Nodes with posterior probability less than 0.95 were collapsed into polytomies, numbers at the tips show the number of genotypes collapsed. Node labels are the point estimates and 95% credible intervals 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.

456 does not resolve relationships among *w*San, *w*Yak, and *w*Tei (Figure 2A). Our phylogram that
457 includes only *D. yakuba*-clade and *w*Mel Wolbachia (based on 643 genes across 644,586 bp)
458 resolves these relationships, placing *w*Tei variants outgroup to sister *w*San and *w*Yak in distinct
459 clades (Figure 2B). We observe 0.0039% third-position pairwise differences between *w*Tei and
460 *w*Yak and between *w*Tei and *w*San, and 0.0017% between *w*San and *w*Yak. The third-position
461 pairwise differences between any of the *D. yakuba*-clade Wolbachia and *w*Mel is 0.11%. For
462 reference, the third-position pairwise difference between *w*Yak and *w*Ri is 2.9%, whereas *w*Ri
463 and *w*Suz differ by only 0.014% (Turelli *et al.* 2018).
464

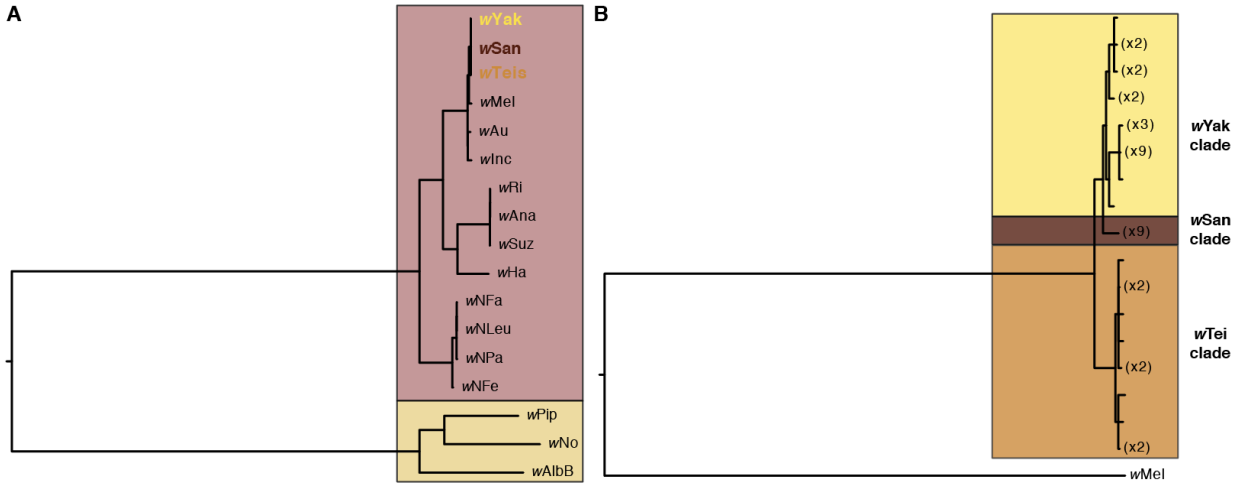

Figure 2 A) Bayesian phylogram placing the *D. yakuba*-clade Wolbachia in the context of other A-, and also B-group, Wolbachia. *w*Pip, wNo, and wAlbB belong to Wolbachia group B, while *D. yakuba* clade-Wolbachia, *w*Mel, *w*Au, *w*Inc, *w*Ri, *w*Ana, *w*Suz, *w*Ha, and Nomada Wolbachia belong to group 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.

465 Our absolute chronogram that includes *w*Mel and *D. yakuba*-clade Wolbachia also indicates
466 three monophyletic groups (Figure 3), whose topology—(*w*Tei, (*w*Yak, *w*San))—agrees with the
467 nuclear topology of the hosts (Turissini and Matute 2017). However, one *D. santomea* line, *D.*
468 *santomea* Quija 37, sampled from the south side of São Tomé (Matute 2010), carries Wolbachia
469 identical to the other eight *w*San samples across 695,118 bp, yet its mtDNA is nested in the *D.*

470 *yakuba* mtDNA cluster (bold blue in Figure 1). This single example of Wolbachia-mtDNA
471 phylogenetic discordance can be explained by either horizontal Wolbachia transfer or paternal
472 transfer of Wolbachia or mtDNA. While rare, paternal Wolbachia transmission has been
473 observed in other Drosophila (Hoffmann and Turelli 1988; Turelli and Hoffmann 1995), as has
474 paternal transmission of mtDNA (Kondo *et al.* 1990). The only other Wolbachia-screened *D.*
475 *santomea* isofemale line with introgressed mtDNA was Wolbachia uninfected.
476

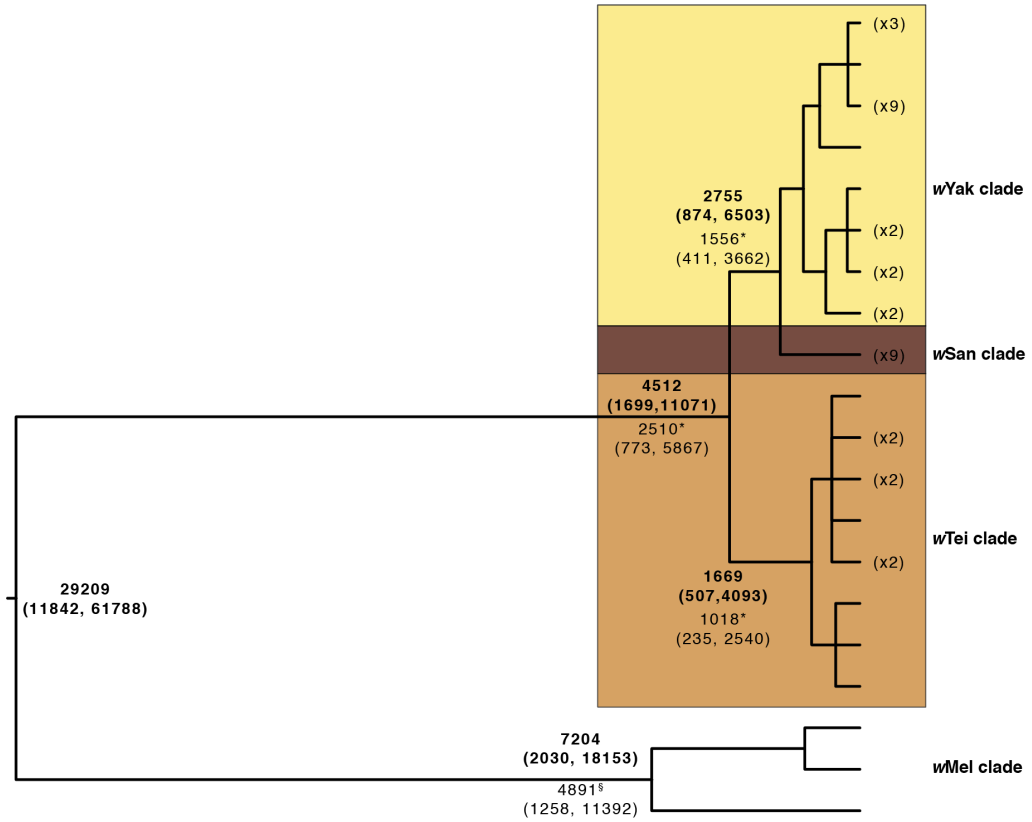

Figure 3 Bayesian chronogram for the *D. yakuba*-clade Wolbachia plus three *w*Mel genomes with absolute age estimates based on the calibration in Richardson *et al*. (2012) in bold. The chronogram was estimated with 621 single-copy genes of identical length in all of the genomes, spanning 624,438 bp. 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).

477 In Figure 3, we present alternative estimates of the divergence times for Wolbachia within
478 and between *D. yakuba*-clade species, within *D. melanogaster*, and between the *D. yakuba* clade
479 and *D. melanogaster.* The estimates from the chronogram that includes all of the indicated
480 Wolbachia variants are in bold. To demonstrate how these time estimates depend on the
481 sequences included in the analyses (through estimates of relative divergence for each codon
482 position), estimates that included only *D. yakuba*-clade Wolbachia or only *w*Mel variants are
483 superimposed and not in bold. These estimates were all obtained using a relaxed-clock analysis
484 with a Γ(2,2) branch-rate prior (results using Γ(7,7) and a strict clock are mentioned below).
485 When considering only the *D. yakuba*-clade Wolbachia (Figure 3), we estimate the root at 2,510
486 years (95% credible interval = 773 to 5,867 years) [our point estimate is 2,642 and credible
487 interval (781, 5,904) using Γ(7,7)] and the *w*Yak-*w*San split at 1,556 years (95% credible interval
488 = 411 to 3,662 years) [point estimate 1,545 and credible interval (401, 3,680) with Γ(7,7)]. The
489 *w*Mel chronogram estimates the most recent common ancestor (MRCA) of *w*Mel at 4,890 years
490 ago (95% credible interval = 1,258 to 11,392 years) [point estimate 4955 and credible interval
491 (1363, 11093) with Γ(7,7)]. Our *w*Mel result is consistent with the Richardson *et al.* (2012)
492 estimate of 8,008 years (95% credible interval (3,263, 13,998). As noted, the alternative branch-
493 rate priors, Γ(2,2) versus Γ(7,7), have little effect on these divergence-times estimates.
494 In contrast, Figure 3 shows that including more divergent variants significantly influences
495 our time estimates. When we analyze the *D. yakuba*-clade and *w*Mel Wolbachia, using the same
496 Γ(2,2) relaxed-clock model, we estimate the root of the *D. yakuba*-clade Wolbachia at 4,512
497 years (95% credible interval = 1,432 to 9,024 years) and the MRCA of *w*Mel at 7,204 years
498 (95% credible interval = 2,511 to 18,739 years), rather than 2,510 and 4,891, respectively.
499 Supplementary Table S4 shows how estimates of the relative amounts of the divergence for the
500 three codon positions vary across the data sets. The underlying model for divergence times
501 assumes constant relative rates at the three positions. Hence, changes in relative rates affect
502 divergence-time estimates.
503 As noted above, our estimates of the time of the MRCA for Wolbachia within the *D. yakuba*
504 clade or within *D. melanogaster* are relatively insensitive to whether the branch-rate prior is
505 Γ(2,2) or Γ(7,7). The estimated divergence time between *w*Mel and the *D. yakuba*-clade
506 Wolbachia is less robust. Using Γ(2,2), we estimate this split at 29,209 years (95% credible
507 interval: 11,842 to 61,788 years); whereas Γ(7,7) produces 49,585 years (17,452 to 106,310

508 years). (A strict clock analysis, corresponding to $\Gamma(n,n)$ as $n \to \infty$, produces 72,612, with 95%
509 credible interval (23,412, 132,276)). An alternative point estimate can be obtained from pairwise
510 differences. Averaging over the *D. yakuba*-clade sequences, we observe an average third-
511 position difference of 0.107% between the *D. yakuba*-clade Wolbachia and *w*Mel, or 0.0535%
512 from tip to root. The "short term evolutionary rate" of divergence within *w*Mel estimated by
513 Richardson *et al.* (2012) produces a point estimate of 78,000 years, which overlaps with the
514 credible interval of our relaxed-tree estimate with a $\Gamma(7,7)$ branch-rate prior. This molecular-
515 clock estimate closely agrees with our strict-clock chronogram analysis. Given that these
516 estimates are much shorter than the divergence time between their reproductively isolated hosts,
517 horizontal Wolbachia transmission must have occurred. As discussed below, the horizontal
518 transmission probably involved intermediate hosts.
519

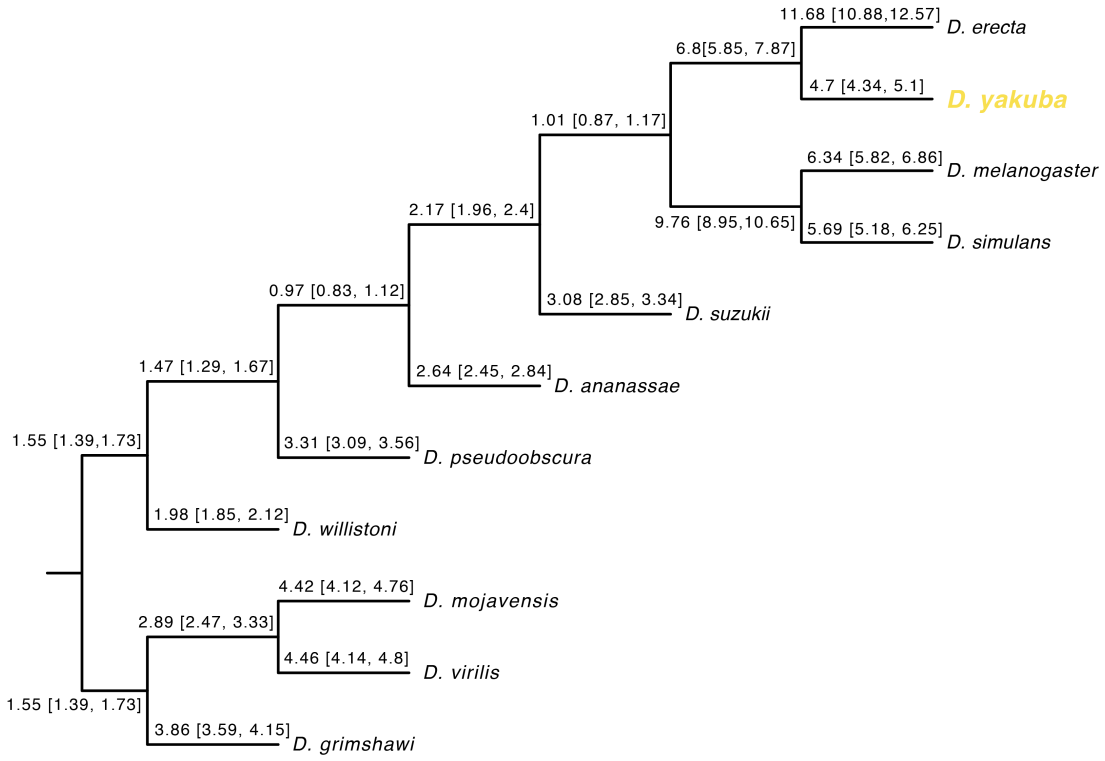

Figure 4 A Bayesian phylogram, with branch lengths fixed to 1, displaying relative rate variation for mtDNA versus nuclear loci. Branches are labeled with quartiles of the estimated third position mtDNA to third position nuclear substitution rate. The nuclear substitution rate was taken as the arithmetic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

520 ***Drosophila*-wide mtDNA to nuclear rate-ratio analysis:** The topologies of the nuclear and
521 mtDNA trees for the 11 reference Drosophila species are completely concordant (Figure 4) and
522 agree with the neighbor-joining result presented in Clark *et al*. (2007). We find the average ratio
523 of mtDNA to nuclear substitution rates across branches vary from about 1× to 10×, consistent
524 with the results of Havird and Sloan (2016) across Diptera.
525
526 **Comparison of mtDNA versus Wolbachia–concordance:** The discordance between the
527 placement of the mtDNA of *D. santomea* line Quija 37, which has mtDNA belonging to the
528 clade associated *D. yakuba* and its Wolbachia, which is clearly *w*San, demonstrates horizontal or
529 paternal transmission. Apart from this, we find no nodes in either the mtDNA or Wolbachia tree
530 with posterior probability at least 0.95 that is inconsistent with a strongly supported node (i.e.,
531 posterior probability at least 0.95) in the other tree. For a more refined test of concordance, we
532 removed the Quija 37 line from our mtDNA and Wolbachia data and computed Bayes Factors.
533 The shared topology model was favored by a Bayes factor of e^{55}, indicating strong support for
534 concordance of the mtDNA and Wolbachia phylogenies.
535
536 **Comparison of mtDNA versus Wolbachia–relative rates:** Following the methods of Turelli *et*
537 *al.* (2018) to look for different mtDNA versus Wolbachia branch lengths, we also found no
538 evidence of variation in the ratio of mtDNA:Wolbachia substitution rates within the *D. yakuba*
539 clade (median 3rd position ratio = 223, quartiles = 186 and 268); the model with all branches
540 sharing the same ratio is favored over the model where each branch has its own ratio by a Bayes
541 factor of e^{7}. For comparison, Turelli *et al.* (2018) found a median ratio of 566 within *D. suzukii*
542 and 406 within *w*Ri-infected *D. ananassae*-subgroup species. These differences in relative rates
543 are broadly compatible with the tenfold variance in relative rates noted above for nuclear versus
544 mitochondrial divergence across Drosophila species. Hence, our data seem compatible with
545 Wolbachia transmission via introgression within the *D. yakuba* clade.
546 Although this result seems to strongly support purely maternal transmission of Wolbachia
547 within species and introgressive transfer between species (with more recent introgression
548 between the sister species *D. yakuba* and *D. santomea*), this interpretation is severely weakened
549 by our "control" analysis that includes *w*Mel-infected *D. melanogaster* (for which introgression
550 is impossible). Including the *D. melanogaster* data, the constant-ratio model was still favored

551 over variable ratios by a Bayes factor of e^{54} (median 3rd position ratio = 297, quartiles = 276 and
552 323). We discuss the implications of this anomalous result below. The key observation is that
553 significantly declining rates of substitution for Wolbachia (and mtDNA, see Ho *et al.* 2005) over
554 time, together with heterogeneity of relative rates (as illustrated in Figure 4), limit the power of
555 relative substitution ratios to differentiate between introgression and horizontal transmission.
556

557 **Analysis of CI factors**

558 In contrast to *w*Mel, which contains only the *cidA-cidB* gene pair (LePage et al. 2017), the *D.*
559 *yakuba*-clade Wolbachia also have both *cidA-cidB*$^{w\text{Yak-clade}}$ and a homologous *cinA-cinB* pair
560 originally identified in group-B *w*Pip (Beckmann *et al*. 2017), which diverged from A-group
561 Wolbachia about 6-46 mya (Werren *et al*. 1995; Meany *et al*. 2018). Among our *w*Yak-clade
562 sequences, there are no single-nucleotide variants within any of the CI-related loci. The *cidB*$^{w\text{Yak-}}$
563 ^{clade} locus has an inversion from amino acids 37–103 relative to the same region in *cidB*$^{w\text{Mel}}$ in
564 every *w*Yak-clade Wolbachia variant we analyzed. This introduces several stop codons, which
565 might render this gene nonfunctional. On the other hand, RNA polymerase should still transcribe
566 a complete polycistronic transcript. Therefore translation of an N-terminally truncated CidB$^{w\text{Yak-}}$
567 ^{clade} protein cannot be ruled out. The sequence differences between *cinA*$^{w\text{Yak-clade}}$ and *cinB*$^{w\text{Yak-clade}}$
568 regions compared to *cinA*$^{w\text{Pip}}$ and *cinB*$^{w\text{Pip}}$ homologs are 1.35% and 0.59%, respectively. In
569 contrast, the average difference between *w*Yak and *w*Pip genomes across all 194 genes (161,655
570 bp) present in single copy and with identical lengths in both genomes is 11.60%. Conversely,
571 outside of the prophage regions, *w*Mel and *w*Yak differ by about 1%. This is consistent with data
572 that indicate WO phage regions have a different evolutionary history than the bulk of the
573 Wolbachia genome (LePage *et al*. 2017; Lindsey *et al*. 2018).

574 The absence of *cin* loci in *w*Mel, combined with the similarity of this region between *w*Yak
575 and distantly related *w*Pip, led us to further evaluate the *cin*-containing regions. Targeted
576 assembly extended the scaffold *w*Yak "702380" containing the *cin* loci (Figure 5) but did not
577 definitively place it relative to the *w*Mel genome. BLAST searches indicated that contig *w*Yak
578 "187383" was likely flanked by *w*Yak "702380". PCR primers designed in both contigs
579 amplified the intervening region (labeled "A" in Figure 5). Subsequent Sanger sequencing
580 revealed that this region contains an ISWpi1 element (Supplemental Information), found in many
581 Wolbachia genomes, but not present at this location in the *w*Mel reference sequence (Cordeaux

582 *et al*. 2008). IS elements encode a transposase gene that mediates their movement (Chandler and
583 Mahillon 2002). ISWpi1 elements are related to the IS5 family, which seems to be restricted to
584 Wolbachia (Duron *et al.* 2005; Cordaux *et al*. 2008). Importantly, ISWpi1 elements occur in
585 more than half of the Wolbachia strains that have been evaluated, including *w*Yak and *w*Mel
586 (Cordaux *et al.* 2008); and these elements occur at variable copy number, potentially facilitating
587 horizontal transmission of ISWpi1 elements and their intervening sequence between Wolbachia
588 variants/strains (Cordaux *et al*. 2008). Following placement of contig *w*Yak "702380" relative to
589 *w*Mel, we aligned these regions and calculated pairwise differences along the chromosome using
590 a sliding window.

591 In addition to containing the *cinA-cinB*, genes absent in *w*Mel, contig *w*Yak "702380" is on
592 average about 10% different from *w*Mel. In contrast, most of *w*Yak is less than 1% different
593 from *w*Mel. Downstream of the *cinA-cinB*$^{w\text{Yak-clade}}$ region, targeted assembly enabled us to join
594 several more contigs. The junctions were corroborated by mapping paired-end reads (Langmead
595 and Salzberg 2012) and visually inspecting the resulting bam files around joined contigs for
596 reads spanning the new junctions and for concordant read-pair mappings. However, our attempts
597 to fully bridge the gap downstream of the *cinA-cinB*$^{w\text{Yak-clade}}$ genes via targeted assembly,
598 scaffolding, and PCR were unsuccessful (see "*w*Yak unassembled gap" in Figure 5). The
599 unassembled region in *w*Yak contains an ISWpi1 element in *w*Mel, Mel #9 (labeled "B" in
600 Figure 5; Cordaux 2008). Although not part of our *w*Yak assembly, homologs of this ISWpi1
601 element appear in assemblies of this region of the WO prophage in several other A-group
602 Wolbachia (Cordaux *et al.* 2008; Bordenstein and Bordenstein 2016), specifically *w*Inc from *D.*
603 *incompta* (Wallau *et al.* 2016) and *w*Ri from *D. simulans* (see Figure 2). In both, we find
604 orthologs to Mel#9 ISWpi1 with corresponding flanking sequence assembled, indicating that this
605 IS element probably occurs in this unassembled region of *w*Yak, *w*San, and *w*Tei. The *w*Yak
606 sequence between these two ISWpi1 elements is highly diverged from *w*Mel in comparison to
607 the rest of the *w*Yak genome (about 10% difference versus 1%). It therefore seems plausible that
608 this region experienced a horizontal transfer event in the ancestor of *w*Yak, *w*Tei, and *w*San,
609 mediated by the flanking ISWpi1 elements.

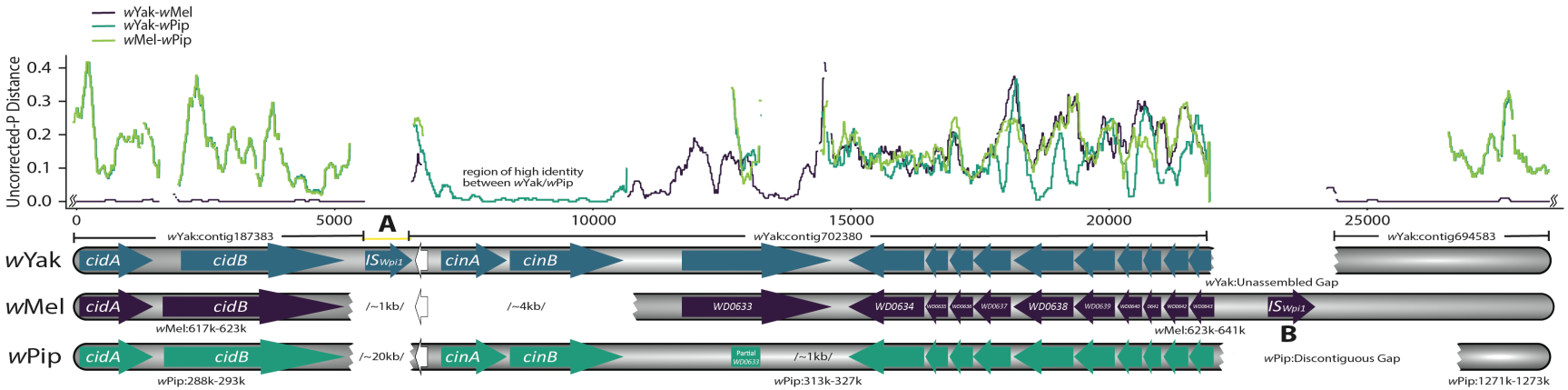

Figure 5 The gene structure and observed, uncorrected, pairwise distances among *w*Yak, *w*Mel, and *w*Pip Wolbachia spanning the region of the WO prophage island where the hypothesized transfer occurred between the ISWpi1 element "A" present in *w*Yak and the ISWpi1 sequence labeled "B" in *w*Mel. Over 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 flanking the region containing the *cinA-cinB* homologs (*w*Yak "187383" and *w*Yak "694583"). 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).

610

611

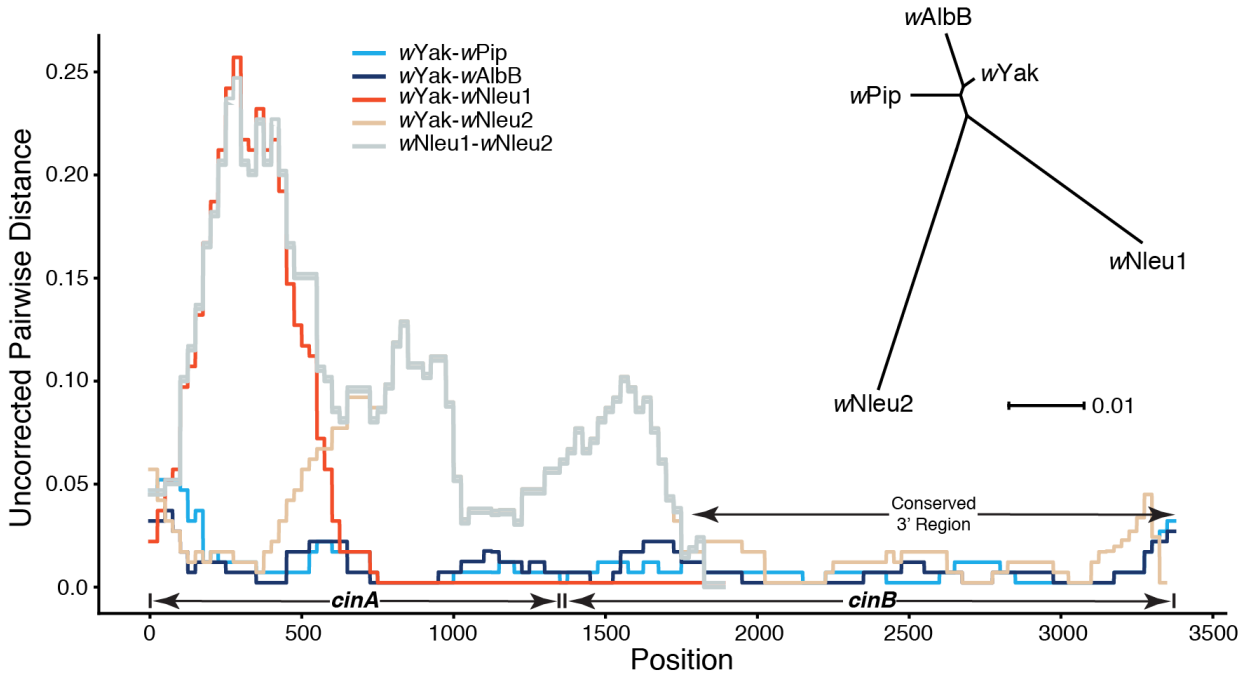

Figure 6 The observed pairwise differences between *cinA* and *cinB* copies in *w*Yak and those in *w*AlbB and *w*Pip, not including flanking regions, are shown. We also plot the pairwise differences between the two copies of these loci found in Nomada Wolbachia (denoted *w*Nleu1 and *w*Nleu2); we chose *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).

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

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

625 The two *cinA-cinB* copies (denoted *w*NLeu1 and *w*NLeu2 in Figure 6) in Nomada
626 Wolbachia are nearly as distinct from each other as they are from the homologs in *w*Pip, *w*AlbB,
627 and *w*Yak (~7% diverged from these strains, Figure 6). However, among the four Wolbachia-
628 infected Nomada species, the orthologs are very similar with *cinA*$^{w\text{NLeu1}}$ having only 0–0.15%
629 pairwise differences among the four strains and *cinA*$^{w\text{NLeu2}}$ having 0–0.56% pairwise differences.
630 (Reconstructions for *cinB* gene copies were more complicated as the *cinB*$^{w\text{NLeu1}}$ copy fails to
631 assemble into a single contig on the 3-prime end.) This pattern suggests that *w*NLeu1 and
632 *w*NLeu2 *cin* copies were acquired by the common ancestor of the four Nomada Wolbachia
633 strains analyzed, followed by cladogenic transfer across Nomada species (Gerth and Bleidorn
634 2016). The highly fragmented assemblies of the four Nomada Wolbachia strains, with duplicate
635 copies confounding assembly, make it difficult to determine the relative positions of the *cinA* and
636 *cinB* copies and if they are likewise flanked by ISWpi1 elements.

637 To determine whether these genes were potentially moved by ISWpi1 elements into the
638 Wolbachia of the *D. yakuba* clade and Nomada, we searched each genome using both the *cinA-*
639 *cinB*$^{w\text{Yak-clade}}$ contig and the flanking ISWpi1 elements. Long repeated elements like ISWpi1
640 (916bp) break most short-read assemblies. Despite this, there is often a small fragment of the
641 element, the length of the short read, on either end of the broken contig, indicative of the repeat
642 element being responsible for terminating contig extension. We looked for the footprint of these
643 elements at the edges of the contigs on which the *cin* genes were found. We found ISWpi1
644 elements in the region flanking the both *cinA*$^{w\text{NLeu1}}$ and *cinA*$^{w\text{NLeu2}}$ copies, consistent with our
645 *w*Yak assembly in which we verified the ISWpi1 element with Sanger sequencing. These data
646 support a role for ISWpi1 in the acquisition of the *cinA-cinB* genes by the Wolbachia in the *D.*

647 *yakuba* clade and the Nomada bees. We conjecture that future work will further implicate
648 ISWpi1 in the horizontal movement of CI loci between Wolbachia.
649
650 **DISCUSSION**
651 Our results indicate introgressive and horizontal Wolbachia acquisition in the *D. yakuba* clade.
652 Our data also indicate horizontal transfer of *cin* CI loci to these Wolbachia (or their A-group
653 ancestors) from highly diverged B-group Wolbachia, with transfer mediated by ISWpi1
654 transposable elements. We discuss these conclusions below.
655
656 **mtDNA Introgression**
657 Our relative mitochondrial chronogram provides strong support for three mtDNA clades,
658 including a monophyletic *D. teissieri* clade that is outgroup to two sister clades: one clade
659 consisting of mtDNA from 14 *D. santomea* individuals, the other contains all *D. yakuba* mtDNA
660 plus mtDNA from 12 *D. santomea* individuals. Our results suggest less mitochondrial
661 introgression in the *D. yakuba* clade than a past report that used sequence data from only two
662 mitochondrial loci (*COII* and *ND5*, 1777 bp) and reported a clade with mtDNA sampled from
663 each species represented (Figure 3 in Bachtrog *et al*. 2006). (Using sequence data from only
664 *COII* and *ND5*, we can replicate this result, indicating data from additional loci are needed to add
665 resolution.) Our results also agree with recent work that demonstrated little nuclear introgression
666 in the *D. yakuba* clade (Turissini and Matute 2017).
667 Our results broadly agree with the results produced by Llopart *et al.* (2014) who assessed
668 mitochondrial introgression between *D. santomea* and *D. yakuba* using whole mitochondrial
669 genomes. They generated a neighbor-joining tree that produced a clade consisting of all *D.*
670 *yakuba* individuals and 10 *D. santomea* individuals, and this clade is sister to a clade with 6 *D.*
671 *santomea* individuals; one *D. santomea* haplotype is outgroup to all other haplotypes included in
672 their analysis (Figure 3 in Llopart *et al.* 2014). Nested within the mixed *D. yakuba* clade, Llopart
673 *et al.* (2014) identified a "hybrid zone clade" that includes *D. yakuba* individuals sampled from
674 São Tomé and from continental Africa and *D. santomea* individuals sampled from both within
675 and outside the Pico de São Tomé hybrid zone (HZ). The sister *D. santomea* clade also contains
676 both HZ and non-HZ individuals. Thus, their analysis and ours provide support for hybridization
677 within and outside of the HZ, leading us to question both the existence of a HZ clade and the

678 claim that these species "share the same mitochondrial genome" (Llopart *et al.* 2014). Instead,
679 both our results and theirs suggest unidirectional introgression of *D. yakuba* mitochondria into *D.*
680 *santomea*; we find 59% of *D. santomea* individuals having *D. yakuba*-like mitochondria and they
681 found 46%.
682 Llopart *et al.* (2014) used a strict molecular clock to estimate the mitochondrial MRCA of *D.*
683 *santomea* and *D. yakuba* at 10,792-17,888 years by calibrating their tree with the *D. yakuba-D.*
684 *erecta* split, estimated at 10.4 mya (Tamura *et al.* 2004). A high level of mitochondrial saturation
685 over time, with an expected value of 1.44 substitutions per synonymous site for the *D. yakuba-D.*
686 *erecta* split, could influence this estimate (Llopart *et al.* 2014). Moreover, Ho *et al.* (2005)
687 demonstrated that the mitochondrial substitution rate resembles an exponential curve, with high
688 short-term substitution rates that approach the mutation rate, then slow to a long-term rate after
689 about 1–2 million years of divergence, far younger than the inferred *D. yakuba-D. erecta* nuclear
690 and mtDNA divergence. Hence, using the slow long-term *D. yakuba-D. erecta* calibration is
691 likely to underestimate the more rapid rates of divergence experienced by *D. yakuba*-clade
692 mtDNA, inflating their short-term divergence-time estimates (Ho *et al*. 2005). If we assume that
693 Wolbachia and mtDNA were transferred by introgression, which our analyses support, our
694 estimates in Figure 3 suggest that *D. santomea* and *D. yakuba* mtDNA diverged more recently,
695 with our point estimates ranging from about 1,500 to 2,800 years.
696
697 **Wolbachia placement, divergence, and acquisition**
698 **Wolbachia placement:** Despite their similarity, *w*San, *w*Tei, and *w*Yak form monophyletic
699 clades with *w*Tei outgroup to sisters *w*San and *w*Yak, recapitulating host relationships (Figure 1;
700 Turissini and Matute 2017). *w*Mel is sister to the *D. yakuba*-clade strains in the A group (Figure
701 2), which also includes *D. simulans* strains (*w*Ri, *w*Au, and *w*Ha), *w*Ana, *w*Suz, and the Nomada
702 Wolbachia (*w*NFa, *w*NLeu, *w*NPa, and *w*NFe). Our phylogram (Figure 2A) places *w*AlbB
703 outgroup to *w*No and *w*Pip strains that diverged from A-group Wolbachia about 6–46 mya
704 (Meany *et al.* 2018).
705
706 **Wolbachia divergence:** Our chronogram analyses (Figure 3) estimate that *D. yakuba*-clade
707 Wolbachia and three *w*Mel variants diverged about 29,000 years ago, and that *w*Tei split from
708 sisters *w*San and *w*Yak about 2,500–4,500 years ago, with *w*San and *w*Yak diverging about

709 1,600-2,800 years ago. We estimate that the two most divergent *w*Mel variants from Richardson
710 *et al.* (2012) and the reference *w*Mel genome split about 4,900–7,200 years ago, indicating more
711 divergence among *w*Mel variants than among *D. yakuba*-clade Wolbachia strains. All of these
712 results depend on the calibration provided by Richardson *et al.* (2012) and the relative accuracy
713 of the underlying models of molecular evolution which assume constant relative rates of change
714 across data partitions. For the deepest divergence in Figure 3, between *D. yakuba*-clade
715 Wolbachia and *w*Mel, we find that the estimated time depends on the variance in our prior for
716 substitution rates across branches, with a strict clock putting the divergence at about 73,000 years
717 rather than 29,000 years obtained with the most variable prior. Despite this uncertainty, the
718 quantitative differences we document for Wolbachia divergence-time estimates do not alter the
719 qualitative conclusion that these Wolbachia did not co-diverge with these hosts, which split
720 several million years ago.
721
722 **Wolbachia acquisition—introgression versus horizontal:** Our divergence-time estimates for
723 the *D. yakuba*-clade Wolbachia versus their hosts preclude cladogenic acquisition. Unlike
724 introgression, horizontal (or paternal) Wolbachia acquisition should produce discordance
725 between phylogenies inferred for Wolbachia and the associated mtDNA. With the notable
726 exception of *D. santomea* line Quija 37, which has mtDNA belonging to the clade associated
727 with *D. yakuba*, but has *w*San Wolbachia, we find no evidence of discordance between the
728 estimated mtDNA and Wolbachia phylogenies. Hence our data indicate that acquisition by
729 introgression is far more common than horizontal transmission between closely related species,
730 consistent with data on acquisition of *w*Ri-like Wolbachia in the *D. melanogaster* species group
731 (Turelli *et al*. 2018). Similarly, consistent with extensive data from *D. melanogaster* (Richardson
732 *et al*. 2012) and *D. simulans* (Turelli *et al*. 2018) and smaller samples from *D. suzukii* and *D.*
733 *ananassae* (Turelli *et al.* 2018), we find only one possible example of paternal transmission or
734 horizontal transmission within *D. yakuba*-clade species.
735 We also investigated an alternative approach to distinguish between introgressive and
736 horizontal Wolbachia acquisition by estimating substitution ratios for mtDNA versus Wolbachia
737 (Turelli *et al.* 2018). Because we could estimate this ratio on each branch, we conjectured that
738 this approach might have greater resolving power than our incompletely resolved mtDNA and
739 Wolbachia phylogenies. We expect higher ratios with horizontal transmission because mtDNA

740 would have been diverging longer than recently transferred Wolbachia. However, this approach
741 assumes that mtDNA and Wolbachia substitution rates remain relatively constant. This is
742 contradicted by the finding of Ho *et al.* (2005) that mtDNA substitution rates decline
743 substantially with increasing divergence time, reaching an asymptote after around 1–2 million
744 years. To calibrate their Wolbachia substitution rate estimates, Richardson *et al.* (2012) used an
745 experimentally observed mitochondrial mutation rate in *D. melanogaster* (6.2×10^{-8} mutations
746 per third-position site per generation) that extrapolates to 62% third-position divergence per
747 million years. This extrapolation is nonsensical for a long-term substitution rate. As summarized
748 by Ho *et al*. (2005), typical mtDNA substitution rates are 0.5–1.5% per coding site per million
749 years. Nevertheless, using the ratio of short-term rates for mtDNA and Wolbachia, Richardson *et*
750 *al.* (2012) produced an estimate of the long-term Wolbachia substitution rate that agrees with
751 independent estimates from Nasonia wasps (Raychoudhury *et al.* 2009) and Nomada bees (Gerth
752 and Bleidorn 2016) derived from much longer divergence times (assuming cladogenic
753 Wolbachia acquisition) (Conner *et al*. 2017). This paradox is resolved if Wolbachia molecular
754 evolution is not subject to the dramatic slowdown in rates seen for mtDNA substitution rates.
755 The apparent difference between mtDNA molecular evolution (dramatic slowdown over
756 longer time scales, Ho *et al*. 2005) and Wolbachia molecular evolution (relative constancy, as
757 inferred from similar rates of differentiation over very different time scales) suggests why our
758 relative-rate test does not reject introgressive transmission of Wolbachia between *D.*
759 *melanogaster* and the *D. yakuba* clade, even though it is clearly impossible. A roughly 50-fold
760 slowdown in mtDNA substitution rates over the time scale of the divergence of *D. melanogaster*
761 from the *D. yakuba* clade, relative to the rate of differentiation within the *D. yakuba* clade,
762 produces comparable mtDNA-Wolbachia substitution ratios for comparisons within the *D.*
763 *yakuba* clade and between *D. melanogaster* and the *D. yakuba* clade. Because of this
764 complication and our conjecture that relative rates of mtDNA versus Wolbachia substitutions
765 over longer periods are likely to mirror the ten-fold differences we see for mtDNA versus
766 nuclear genes, phylogenetic discordance between mtDNA and Wolbachia is clearly a much more
767 robust indicator of horizontal (or paternal) Wolbachia transmission. Nevertheless, additional
768 examples of cladogenic Wolbachia acquisition are needed to better understand relative rates and
769 patterns of Wolbachia, mtDNA and nuclear differentiation over different time scales.

770 Our divergence-times estimate of the *D. yakuba*-clade Wolbachia versus their hosts precludes
771 cladogenic transmission; and our phylogenetic analyses suggest that these species share very
772 similar Wolbachia because of introgression, as originally argued by Lachaise *et al.* (2000).
773 However, under either introgressive or horizontal transfer of Wolbachia, we expect the donor
774 species Wolbachia sequences would appear paraphyletic when analyzed jointly with the
775 Wolbachia from the recipient. Paraphyly allowed Turelli *et al*. (2018) to infer that *D. simulans*
776 likely obtained its Wolbachia from *D. ananassae* and *D. subpulchrella* likely obtained its
777 Wolbachia from *D. suzukii*. Paraphyly is generally expected soon after gene flow stops between
778 populations. As noted by Hudson and Coyne (2002, Fig. 1), the time scale expected to produce
779 reciprocal monophyly for mtDNA (and Wolbachia) under a neutral model of molecular evolution
780 is on the order of the effective size of the species. Our results in Figure 3 indicate that, at least for
781 our small samples, reciprocal monophyly for the Wolbachia in these three species has been
782 achieved within a few thousand years. This suggests that reciprocal monophyly has been
783 accelerated by species-specific selective sweeps within the Wolbachia or mtDNA of these
784 species. This conjecture may be testable from estimates of host fitness using transinfected versus
785 native Wolbachia.
786
787 **Patterns of CI and Wolbachia genome evolution**
788 Wolbachia in all three *D. yakuba*-clade hosts cause both intra-and interspecific CI (Cooper *et al.*
789 2017), despite originally being characterized as non-CI causing (Zabalou *et al.* 2004; Charlat *et*
790 *al.* 2004). *D. yakuba*-clade Wolbachia CI is relatively weak, and its strength can vary subtly
791 among *w*Tei variants and *D. teissieri* backgrounds (Table 3 in Cooper *et al.* 2017). Differences in
792 CI among Wolbachia variants has also been demonstrated in interspecific backgrounds where
793 *w*Tei caused stronger CI in a *D. simulans* background (97.2 ± 1.3 SE percent embryo mortality)
794 than either *w*Yak (26.5 ± 4.2 SE, percent embryo mortality) or *w*San (24.0 ± 4.1 SE, percent
795 embryo mortality) (Zabalou et al. 2008). Both *w*Yak and *w*San induced CI in *D. simulans*
796 comparable in intensity to that found by Cooper *et al.* (2017, Figure 3) in their original hosts.
797 Surprisingly, Zabalou *et al*. (2008) found that the strength of CI induced by *w*Tei in *D. simulans*
798 even eclipsed that of *w*Ri (89.8 ± 4.5 SE percent embryo mortality). These results are hard to
799 reconcile with the fact that WO phage loci known to underlie CI do not vary within or among *D.*
800 *yakuba*-clade Wolbachia variants we examined. The nearly complete CI induced by *w*Tei in *D.*

801 *simulans* may depend on CI-causing factors yet to be identified or differences in gene
802 expression.
803 In each *D. yakuba*-clade Wolbachia variant included in our analyses, we find a disruption of
804 *cidB*$^{w\text{Yak-clade}}$ with an inversion from amino acids 37–103 relative to the same region in sister
805 *w*Mel. The inversion introduces multiple stop codons that could render this gene nonfunctional.
806 Fixation of loss-of-function mutations in CI-causing loci is consistent with theoretical analyses
807 showing that selection on Wolbachia within host lineages does not act to increase or maintain CI
808 (Prout 1994; Turelli 1994; Haygood and Turelli 2009); indeed, we have also recently observed a
809 single mutation that disrupts *cidB* in non-CI causing *w*Mau Wolbachia that infect *D. mauritiana*
810 on the island of Mauritius (Meany et al. 2018). We expect that future genomic analyses will
811 produce additional examples.
812 All *D. yakuba*-clade Wolbachia genomes included in our analysis harbor *cinA-cinB* loci
813 originally discovered in the *w*Pip strain that diverged from A-group Wolbachia, including the *D.*
814 *yakuba*-clade variants, about 6–46 mya (Meany *et al*. 2018). *cin* loci are also present in B-group
815 *w*AlbB that infects *Ae. albopictus* and in A-group *w*NFe, *w*NPa, *w*NLeu, and *w*NFa Wolbachia
816 that infect Nomada bees. *cin* loci are absent from *w*Mel, but the wYak contig containing these
817 loci is about 10% diverged from *w*Mel, while observed divergence between wYak and wMel
818 across the rest of the genome is less than 1%. The *w*Yak-clade *cin* loci share about 97%
819 similarity with the divergent B-group *w*Pip strain. *w*Yak-clade *cin* loci are more similar to *cinA-*
820 *cinB* from the B-group Wolbachia *w*Pip and *w*AlbB than to those in A-group Nomada Wolbachia
821 strains, which have two sets of *cin* loci that are as divergent from each other as they are from
822 these regions in *w*Yak and in B-group Wolbachia. These observations suggest independent
823 horizontal transfer of *cin* loci into *w*Yak and Nomada Wolbachia.
824 We hypothesize that repeat regions constructed of ISWpi1 element paralogs near WO-phage
825 integrations may serve to move CI loci as transposon cargo via the excision of flanking ISWpi1
826 elements, followed by homologous recombination within the elements, independently of phages.
827 Horizontal Wolbachia acquisition is common in Drosophila (Turelli *et al.* 2018) and other
828 species (O'Neill *et al.* 1992), suggesting that double infections, which could provide the
829 opportunity for ISWpi1 mediated transfer of *cin* loci, must also be common, even if transient. (A
830 second infection need not become vertically transmitted for horizontal gene transfer via ISWpi1
831 elements to occur.) In contrast, phage particles or virions could be introduced by a vector and

832 provide the opportunity for ISWpi1 mediated transfer (Ahmed *et al.* 2015; Brown and Lloyd
833 2015), without the presence of a double Wolbachia infection. Determining whether the insertion
834 of these loci was derived from a prophage region of the Wolbachia genome, or from a phage
835 genome encapsulated in a phage particle, remains an open question. While the ISWpi1 element
836 in *w*Mel (Mel #9 labeled "B" in Figure 5; Cordaux 2008) is not part of our *w*Yak assembly,
837 homologs of this element are present in assemblies of several other A-group Wolbachia
838 including *w*Inc and *w*Ri (Cordaux *et al.* 2008; Bordenstein and Bordenstein 2016). We predict
839 this element occurs in the unassembled region of our *w*Yak assembly, but long-read sequencing
840 will be required to test this hypothesis. Footprints of ISWpi1 elements in the region flanking the
841 *cinA* genes for both copies of the gene in the Nomada Wolbachia provide further support for our
842 hypothesis. Future efforts to generate and analyze long-read Wolbachia assemblies will help
843 confirm the role of ISWpi1 elements in horizontal transfer of CI loci. Ultimately, understanding
844 the ecology of horizontal Wolbachia transmission is crucial to explain mechanisms of horizontal
845 Wolbachia acquisition; horizontal Wolbachia acquisition, even if ephemeral, provides an
846 opportunity for horizontal transfer of CI loci between Wolbachia variants/strains via ISWpi1
847 elements.
848

849 **ACKNOWLEDGMENTS**

850 We thank Tim Wheeler for help in the laboratory and for taking pictures of host species. John
851 Beckmann and Amelia Lindsey provided comments that were very helpful. Research reported in
852 this publication was supported by the National Institute Of General Medical Sciences of the
853 National Institutes of Health (NIH) under Award Number R35GM124701 to B.S.C,
854 R01GM121750 to D.R.M, and R01GM104325 to M.T. The content is solely the responsibility of
855 the authors and does not necessarily represent the official views of the NIH.

856 **LITERATURE CITED**

857 Ahmed, M. Z., S.-J. Li, X. Xue, X.-J. Yin, S.-X. Ren *et al.*, 2015 The Intracellular Bacterium
858 Wolbachia Uses Parasitoid Wasps as Phoretic Vectors for Efficient Horizontal
859 Transmission. PLoS Pathogens 11**:** e1004672.
860 Azad, A. F., and C. B. Beard, 1998 Rickettsial pathogens and their arthropod vectors. Emerging
861 Infectious Diseases 4**:** 179-186.
862 Bachtrog, D., K. Thornton, A. Clark and P. Andolfatto, 2006 Extensive introgression of
863 mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group.
864 Evolution 60**:** 292-302.
865 Bakovic V., M. Schebeck, A. Telschow, C. Stauffer, H. Schuler, 2018 Spatial spread
866 of Wolbachia in *Rhagoletis cerasi* populations. Biol. Lett. 14:20180161
867 Baldo, L., N. A. Ayoub, C. Y. Hayashi, J. A. Russell, J. K. Stahlhut *et al.*, 2008 Insight into the
868 routes of Wolbachia invasion: high levels of horizontal transfer in the spider genus
869 Agelenopsis revealed by Wolbachia strain and mitochondrial DNA diversity. Mol. Ecol.
870 17**:** 557-569.
871 Baldo, L., J. C. D. Hotopp, K. A. Jolley, S. R. Bordenstein, S. A. Biber *et al.*, 2006 Multilocus
872 sequence typing system for the endosymbiont *Wolbachia pipientis*. Appl. Environ.
873 Microbiol. 72**:** 7098-7110.
874 Baldridge, G. D., N. Y. Burkhardt, J. A. Simser, T. J. Kurtti and U. G. Munderloh, 2004
875 Sequence and Expression Analysis of the *ompA* Gene of *Rickettsia peacockii*, an
876 Endosymbiont of the Rocky Mountain Wood Tick, *Dermacentor andersoni*. Appl.
877 Environ. Microbiol. 70**:** 6628.
878 Bandi, C., T. J. Anderson, C. Genchi and M. L. Blaxter, 1998 Phylogeny of Wolbachia in filarial
879 nematodes. Proc. Biol. Sci. 265**:** 2407-2413.
880 Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin *et al.*, 2012 SPAdes: a new
881 genome assembly algorithm and its applications to single-cell sequencing. J. Comput.
882 Biol. 19**:** 455-477.
883 Beckmann, J. F., M. Bonneau, H. Chen, M. Hochstrasser, D. Poinsot *et al.*, The Toxin &
884 Antidote Model of Cytoplasmic Incompatibility: Genetics and Evolutionary Implications.
885 Trends Genet. In press.
886 Beckmann, J. F., and A. M. Fallon, 2013 Detection of the Wolbachia protein WPIP0282 in
887 mosquito spermathecae: implications for cytoplasmic incompatibility. Insect Biochem.
888 Mol. Biol. 43**:** 867-878.
889 Beckmann, J. F., J. A. Ronau and M. Hochstrasser, 2017 A Wolbachia deubiquitylating enzyme
890 induces cytoplasmic incompatibility. Nat. Microbiol. 2: 17007.
891 Bordenstein, S. R., and S. R. Bordenstein, 2016 Eukaryotic association module in phage WO
892 genomes from Wolbachia. Nat. Commun. 7**:** 13155.
893 Bouchon, D., T. Rigaud and P. Juchault, 1998 Evidence for widespread Wolbachia infection in
894 isopod crustaceans: molecular identification and host feminization. Proc. Biol. Sci. 265:
895 1081-1090.
896 Braig, H. R., W. Zhou, S. L. Dobson and S. L. O'Neill, 1998 Cloning and characterization of a
897 gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia*
898 *pipientis*. J. Bacteriol. 180: 2373-2378.
899 Brand, C. L., S. B. Kingan, L. Wu and D. Garrigan, 2013 A Selective Sweep across Species
900 Boundaries in Drosophila. Mol. Biol. Evol. 30: 2177-2186.

Brown, A. N., and V. K. Lloyd, 2015 Evidence for horizontal transfer of Wolbachia by a Drosophila mite. Exp. Appl. Acarol. 66: 301-311.

Brownlie, J. C., B. N. Cass, M. Riegler, J. J. Witsenburg, I. Iturbe-Ormaetxe *et al.*, 2009 Evidence for Metabolic Provisioning by a Common Invertebrate Endosymbiont, *Wolbachia pipientis*, during Periods of Nutritional Stress. PLoS Pathogens 5**:** e1000368.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: 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.

Frydman, H. M., J. M. Li, D. N. Robson and E. Wieschaus, 2006 Somatic stem cell niche tropism in Wolbachia. Nature 441**:** 509-512.

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.

Gibson, K. E., Y. Rikihisa, C. Zhang and C. Martin, 2005 *Neorickettsia risticii* is vertically transmitted in the trematode *Acanthatrium oregonense* and horizontally transmitted to bats. Environ. Microbiol. 7**:** 203-212.

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.

991 Hoffmann, A. A., and M. Turelli, 1997 Cytoplasmic incompatibility in insects, pp. 42–80 in
992 Influential Passengers: Inherited Microorganisms and Arthropod Reproduction, edited by
993 S. L. O'Neill, A. A. Hoffmann, and J. H. Werren. Oxford University Press, New York.
994 Hoffmann, A. A., M. Turelli and G. M. Simmons, 1986 Unidirectional incompatibility between
995 populations of *Drosophila simulans*. Evolution 40**:** 692-701.
996 Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot *et al.*, 2016 RevBayes: Bayesian
997 Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification
998 Language. Syst. Biol. 65: 726-736.
999 Hoskins, R. A., J. W. Carlson, K. H. Wan, S. Park, I. Mendez *et al.*, 2015 The Release 6
1000 reference sequence of the *Drosophila melanogaster* genome. Genome Res. 25**:** 445-458.
1001 Hudson, R. R. and J. A. Coyne, 2002 Mathematical consequences of the genealogical species
1002 concept. Evolution 56:1557–1565.
1003 Huigens, M. E., R. P. de Almeida, P. A. Boons, R. F. Luck and R. Stouthamer, 2004 Natural
1004 interspecific and intraspecific horizontal transfer of parthenogenesis-inducing Wolbachia
1005 in Trichogramma wasps. Proc. Biol. Sci. 271**:** 509-515.
1006 Huigens, M. E., R. F. Luck, R. H. Klaassen, M. F. Maas, M. J. Timmermans *et al.*, 2000
1007 Infectious parthenogenesis. Nature 405**:** 178-179.
1008 Hunter, S. S., R. T. Lyon, B. A. J. Sarver, K. Hardwick, L. J. Forney *et al.*, 2015 Assembly by
1009 Reduced Complexity (ARC): a hybrid approach for targeted assembly of homologous
1010 sequences. bioRxiv**:** 014662.
1011 Hurst, G. D. D., and F. M. Jiggins, 2000 Male-killing bacteria in insects: mechanisms, incidence,
1012 and implications. Emerg. Infect. Dis. 6**:** 329-336.
1013 Jackman, S. D., B. P. Vandervalk, H. Mohamadi, J. Chu, S. Yeo *et al.*, 2017 ABySS 2.0:
1014 resource-efficient assembly of large genomes using a Bloom filter. Genome Res. 27**:** 768-
1015 777.
1016 Jaenike, J., K. A. Dyer, C. Cornish and M. S. Minhas, 2006 Asymmetrical reinforcement and
1017 Wolbachia infection in Drosophila. PLoS Biol. 4**:** 1852-1862.
1018 Joshi, N. A., and J. N. Fass, 2011 Sickle: A sliding-window, adaptive, quality-based trimming
1019 tool for FastQ files, pp.
1020 Kaltz, O., and J. C. Koella, 2003 Host growth conditions regulate the plasticity of horizontal and
1021 vertical transmission in *Holospora undulata*, a bacterial parasite of the protozoan
1022 *Paramecium caudatum*. Evolution 57**:** 1535-1542.
1023 Kamimura, Y., and H. Mitsumoto, 2012 Lock-and-key structural isolation between sibling
1024 Drosophila species. Entomol. Sci. 15**:** 197-201.
1025 Katoh, K., and D. M. Standley, 2013 MAFFT Multiple Sequence Alignment Software Version 7:
1026 Improvements in Performance and Usability. Mol. Biol. Evol. 30**:** 772-780.
1027 Kern, A. D., D. A. Barbash, J. Chang Mell, A. Jensen and D. Hupalo, 2015 Highly Constrained
1028 Intergenic Drosophila Ultraconserved Elements Are Candidate ncRNAs. Genome Biol.
1029 Evol. 7**:** 689-698.
1030 Klasson, L., T. Walker, M. Sebaihia, M. J. Sanders, M. A. Quail *et al.*, 2008 Genome evolution
1031 of Wolbachia strain *w*Pip from the *Culex pipiens* group. Mol. Biol. Evol. 25**:** 1877-1887.
1032 Klasson, L., J. Westberg, P. Sapountzis, K. Näslund, Y. Lutnaes *et al.*, 2009 The mosaic genome
1033 structure of the Wolbachia *w*Ri strain infecting *Drosophila simulans*. PNAS 106**:** 5725-
1034 5730.
1035 Kondo, R., Y. Satta, E. T. Matsuura, H. Ishiwa, N. Takahata *et al.*, 1990 Incomplete maternal
1036 transmission of mitochondrial DNA in Drosophila. Genetics 126**:** 657-663.

1037 Kriesner, P., W. R. Conner, A. R. Weeks, M. Turelli and A. A. Hoffmann, 2016 Persistence of a
1038 Wolbachia infection frequency cline in *Drosophila melanogaster* and the possible role of
1039 reproductive dormancy. Evolution 70**:** 979-997.
1040 Kriesner, P., A. A. Hoffmann, S. F. Lee, M. Turelli and A. R. Weeks, 2013 Rapid Sequential
1041 Spread of Two Wolbachia Variants in *Drosophila simulans*. PLoS Path. 9.
1042 Kulathinal, R. J., L. S. Stevison and M. A. F. Noor, 2009 The Genomics of Speciation in
1043 Drosophila: Diversity, Divergence, and Introgression Estimated Using Low-Coverage
1044 Genome Sequencing. PLoS Genet. 5**:** e1000550.
1045 Lachaise, D., M. Harry, M. Solignac, F. Lemeunier, V. Benassi *et al.*, 2000 Evolutionary
1046 novelties in islands: *Drosophila santomea*, a new *melanogaster* sister species from São
1047 Tomé. Proc. Sci. Biol. 267**:** 1487-1495.
1048 Langley, C. H., and W. M. Fitch, 1974 An examination of the constancy of the rate of molecular
1049 evolution. J. Mol. Evol. 3**:** 161-177.
1050 Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nature
1051 Methods 9**:** 357-359.
1052 Laven, H., 1951 Crossing Experiments with Culex Strains. Evolution 5**:** 370-375.
1053 Laven, H., 1967 Eradication of *Culex pipiens fatigans* through cytoplasmic incompatibility.
1054 Nature 216**:** 383-384.
1055 LePage, D. P., J. A. Metcalf, S. R. Bordenstein, J. On, J. I. Perlmutter *et al.*, 2017 Prophage WO
1056 genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility. Nature
1057 543**:** 243-247.
1058 Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping
1059 and population genetical parameter estimation from sequencing data. Bioinformatics 27**:**
1060 2987-2993.
1061 Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler
1062 transform. Bioinformatics 25**:** 1754-1760.
1063 Li, S.-J., M. Z. Ahmed, N. Lv, P.-Q. Shi, X.-M. Wang *et al.*, 2017 Plantmediated horizontal
1064 transmission of Wolbachia between whiteflies. ISME Journal 11**:** 1019-1028.
1065 Lindsey, A. R. I., D. W. Rice, S. R. Bordenstein, A. W. Brooks, S. R. Bordenstein *et al.*, 2018
1066 Evolutionary Genetics of Cytoplasmic Incompatibility Genes *cifA* and *cifB* in Prophage
1067 WO of *Wolbachia*. Genome Biol. Evol. 10**:** 434-451.
1068 Llopart, A., D. Herrig, E. Brud and Z. Stecklein, 2014 Sequential adaptive introgression of the
1069 mitochondrial genome in *Drosophila yakuba* and *Drosophila santomea*. Mol. Ecol. 23**:**
1070 1124-1136.
1071 Lohse, K., M. Clarke, M. G. Ritchie and W. J. Etges, 2015 Genome-wide tests for introgression
1072 between cactophilic Drosophila implicate a role of inversions during speciation.
1073 Evolution 69**:** 1178-1190.
1074 Long, S. W., X. Zhang, J. Zhang, R. P. Ruble, P. Teel *et al.*, 2003 Evaluation of transovarial
1075 transmission and transmissibility of *Ehrlichia chaffeensis* (Rickettsiales:
1076 Anaplasmataceae) in *Amblyomma americanum* (Acari: Ixodidae). J. Med. Entomol. 40**:**
1077 1000-1004.
1078 Martinez, J., B. Longdon, S. Bauer, Y.-S. Chan, W. J. Miller *et al.*, 2014 Symbionts Commonly
1079 Provide Broad Spectrum Resistance to Viruses in Insects: A Comparative Analysis of
1080 Wolbachia Strains. PLoS Path. 10**:** e1004369.
1081 Matute, D. R., 2010 Reinforcement of gametic isolation in Drosophila. PLoS Biol. 8**:** e1000341.

1082 Matute, D. R., and J. F. Ayroles, 2014 Hybridization occurs between *Drosophila simulans* and
1083 *D. sechellia* in the Seychelles archipelago. J. Evol. Biol. 27**:** 1057-1068.
1084 Matute, D. R., and J. A. Coyne, 2010 Intrinsic reproductive isolation between two sister species
1085 of Drosophila. Evolution 64**:** 903-920.
1086 Mavingui, P., C. Valiente Moro, V. Tran-Van, F. Wisniewski-Dye, V. Raquin *et al.*, 2012
1087 Whole-genome sequence of Wolbachia strain *w*AlbB, an endosymbiont of tiger mosquito
1088 vector *Aedes albopictus*. J. Bacteriol. 194**:** 1840.
1089 McMeniman, C. J., R. V. Lane, B. N. Cass, A. W. C. Fong, M. Sidhu *et al.*, 2009 Stable
1090 introduction of a life-shortening Wolbachia infection into the mosquito *Aedes aegypti*.
1091 Science 323**:** 141-144.
1092 Meany, M. K., W. R. Conner, S. V. Richter, J. A. Bailey, M. Turelli *et al.*, 2018 Loss of
1093 cytoplasmic incompatibility and minimal fecundity effects explain relatively low
1094 Wolbachia frequencies in *Drosophila mauritiana*. bioRxiv**:** 461574.
1095 Moriyama, M., N. Nikoh, T. Hosokawa and T. Fukatsu, 2015 Riboflavin Provisioning Underlies
1096 Wolbachia's Fitness Contribution to Its Insect Host. mBio 6. e01732-15.
1097 Munderloh, U. G., and T. J. Kurtti, 1995 Cellular and molecular interrelationships between ticks
1098 and prokaryotic tick-borne pathogens. Annu. Rev. Entomol. 40**:** 221-243.
1099 Nosil, P., 2013 Degree of sympatry affects reinforcement in Drosophila. Evolution 67**:** 868-872.
1100 Nurk, S., A. Bankevich, D. Antipov, A. Gurevich, A. Korobeynikov *et al.*, 2013 Assembling
1101 Genomes and Mini-metagenomes from Highly Chimeric Reads, pp. 158-170 in *Research*
1102 *in Computational Molecular Biology*, edited by M. Deng, R. Jiang, F. Sun and X. Zhang.
1103 Springer Berlin Heidelberg, Berlin, Heidelberg.
1104 O'Connor, L., C. Plichart, A. C. Sang, C. L. Brelsfoard, H. C. Bossin *et al.*, 2012 Open release of
1105 male mosquitoes infected with a Wolbachia biopesticide: field performance and infection
1106 containment. PLoS Negl. Trop. Dis. 6**:** e1797.
1107 O'Neill, S. L., R. Giordano, A. M. Colbert, T. L. Karr and H. M. Robertson, 1992 16S rRNA
1108 phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic
1109 incompatibility in insects. PNAS 89**:** 2699-2702.
1110 O'Neill, S. L., A. A. Hoffman and J. H. Werren, 1998 *Influential Passengers: Inherited*
1111 *Microorganisms and Arthropod Reproduction*. Oxford University Press.
1112 Perlman Steve, J., S. Hunter Martha and E. Zchori-Fein, 2006 The emerging diversity of
1113 Rickettsia. Proc. Sci. Biol. 273**:** 2097-2106.
1114 Pietri, J. E., H. DeBruhl and W. Sullivan, 2016 The rich somatic life of Wolbachia. Microbiol.
1115 Open 5**:** 923-936.
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.

Vavre, F., L. Mouton and B. A. Pannebakker, 2009 Chapter 12 *Drosophila*–Parasitoid Communities as Model Systems for Host–*Wolbachia* Interactions, pp. 299-331 in *Advances in Parasitology*. Academic Press, San Diego, California.

Wallau, G. L., M. T. da Rosa, F. C. De Re and E. L. Loreto, 2016 Wolbachia from *Drosophila incompta*: just a hitchhiker shared by Drosophila in the New and Old World? Insect Mol. Biol. 25**:** 487-499.

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

Weinert, L. A., E. V. Araujo-Jnr, M. Z. Ahmed and J. J. Welch, 2015 The incidence of bacterial endosymbionts in terrestrial arthropods. Proc. Biol. Sci. 282**:** 20150249.

Werren, J. H., and D. M. Windsor, 2000 Wolbachia infection frequencies in insects: evidence of a global equilibrium? Proc. Biol. Sci. 267**:** 1277-1285.

Werren, J. H., W. Zhang and L. R. Guo, 1995 Evolution and phylogeny of Wolbachia: reproductive parasites of arthropods. Proc. Biol. Sci. 261**:** 55-63.

Wu, M., L. V. Sun, J. Vamathevan, M. Riegler, R. Deboy *et al.*, 2004 Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. PLoS Biol. 2**:** e69.

Yen, J. H., and A. R. Barr, 1971 New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens L*. Nature 232**:** 657-658.

Yukilevich, R., 2012 Asymmetrical patterns of speciation uniquely support reinforcement in Drosophila. Evolution 66**:** 1430-1446.

Zabalou, S., A. Apostolaki, S. Pattas, Z. Veneti, C. Paraskevopoulos *et al.*, 2008 Multiple rescue factors within a Wolbachia strain. Genetics 178**:** 2145-2160.

Zabalou, S., S. Charlat, A. Nirgianaki, D. Lachaise, H. Merçot *et al.*, 2004 Natural Wolbachia infections in the *Drosophila yakuba* species complex do not induce cytoplasmic incompatibility but fully rescue the *w*Ri modification. Genetics 167**:** 827-834.

1211 **SUPPLEMENTAL INFORMATION**

1212

1213 **SUPPLEMENTAL MATERIALS AND METHODS**

1214 **PCR amplification and Sanger Sequencing of "A" and "B" regions in Figure 5.**

1215

1216 DNA was extracted from each isofemale (yakuba_S09_L24, yakuba_S01_L19, and
1217 yakuba_S01_L31) line using a standard 'squish' buffer protocol (Gloor *et al*. 1993). A standard
1218 PCR assay (Simpliamp ThermoCycler, Applied Biosystems, Singapore) was performed using the
1219 GoTaq™ DNA Polymerase (Promega™ , Wisconsin, USA) master mix. PCR program began
1220 with 3 minutes at 94°C, followed by 34 cycles of 30 seconds at 94°C, 30 seconds at 55°C, and 1
1221 minute and 15 seconds at 72°C. The profile finished with 8 minutes at 72°C. The amplified
1222 PCR products were visualized on a 1% agarose gel that included a 1kb ladder.

1223 For each fly line, each of the three yak_ISWpi1 primer sets were assayed (Table S1). CinA
1224 primer set consistently returned a 100-150 bp product, 2CinA returned a ~1000 bp product, and
1225 692607_694583 primer set returned no products. Touchdown PCR with an annealing gradient
1226 from 70°C to 55°C was performed with the same results to the standard PCR run from above.

1227 For each fly line, both Mel_ISWPi1 primer sets were assayed (Table S1). All assays
1228 returned no products. Touchdown PCR with an annealing gradient from 70°C to 55°C was
1229 performed and also returned no products. Annealing temp was lowered from 55°C to 50°C and
1230 the PCR was repeated, and again no products were returned.

1231 The ~1000 bp PCR product from yakuba_S09_L24 was cleaned using the QiaQuick PCR
1232 purification kit (Qiagen, Germany). The cleaned PCR product was sent to Eurofins (Kentucky)
1233 for Sanger sequencing.

1234
1235
1236
1237
1238
1239
1240
1241
1242

1243 **SUPPLEMENTAL TABLES**

1244

1245 **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

1246

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

1247

1248

1249

1250

1251 **Table S3.** Near-universal, single-copy proteobacteria genes used to assess assembly
1252 completeness. We used BUSCO v. 3.0.0 to identify complete and fragmented proteobacteria
1253 genes (out of 221) found in our draft Wolbachia genomes. For comparison, we provide results
1254 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

1255

1256

1257

1258

1259

1260

1261

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})