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

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

Michael Turelli[†]

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

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

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

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

Running title: Transfer of *Wolbachia* and CI loci

Key words: cytoplasmic incompatibility, horizontal gene transfer, hybridization, transposable elements, WO phage

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

ABSTRACT

1   Maternally transmitted *Wolbachia* infect about half of insect species, yet the predominant

2   mode(s) of *Wolbachia* acquisition remains uncertain. Species-specific associations could be old,

3   with *Wolbachia* and hosts co-diversifying (i.e., cladogenic acquisition), or relatively young and

4   acquired by horizontal transfer or introgression. The three *Drosophila yakuba*-clade hosts ((*D.*

5   *santomea, D. yakuba*), *D. teissieri*) diverged about three million years ago and currently

6   hybridize on Bioko and São Tomé, west African islands. Each species is polymorphic for nearly

7   identical *Wolbachia* that cause weak cytoplasmic incompatibility (CI)–reduced egg hatch when

8   uninfected females mate with infected males. *D. yakuba*-clade *Wolbachia* are closely related to

9   *w*Mel, globally polymorphic in *D. melanogaster*. We use draft *Wolbachia* and mitochondrial

10 genomes to demonstrate that *D. yakuba*-clade phylogenies for *Wolbachia* and mitochondria tend

11 to follow host nuclear phylogenies. However, roughly half of *D. santomea* individuals, sampled

12 both inside and outside of the São Tomé hybrid zone, have introgressed *D. yakuba* mitochondria.

13 Both mitochondria and *Wolbachia* possess far more recent common ancestors than the bulk of

14 the host nuclear genomes, precluding cladogenic *Wolbachia* acquisition. General concordance of

15 *Wolbachia* and mitochondrial phylogenies suggests that horizontal transmission is rare, but

16 varying relative rates of molecular divergence complicate chronogram-based statistical tests.

17 Loci that cause CI in *w*Mel are disrupted in *D. yakuba*-clade *Wolbachia*; but, a second set of loci

18 predicted to cause CI are located in the same WO prophage region. These alternative CI loci

19 seem to have been acquired horizontally from distantly related *Wolbachia*, with transfer

20 mediated by flanking *Wolbachia*-specific ISWpi1 transposons.

21

22

23

24

25

26

27

28

29

INTRODUCTION

Endosymbiotic *Wolbachia* bacteria infect many arthropods (Bouchon *et al.* 1998; Hilgenboecker *et al.* 2008), including about half of all insect species (Werren and Windsor 2000; Weinert *et al.* 2015). *Wolbachia* often manipulate host reproduction, facilitating spread to high frequencies within host species (Laven 1951; Yen and Barr 1971; Turelli and Hoffmann 1991; Rousset *et al.* 1992; O'Neill *et al.* 1998; Weeks *et al.* 2007; Kriesner *et al.* 2016; Turelli *et al*. 2018). In *Drosophila*, reproductive manipulations include cytoplasmic incompatibility (CI) and male killing (Hoffmann *et al.* 1986; Hoffmann and Turelli 1997; Hurst and Jiggins 2000). CI reduces the egg hatch of uninfected females mated with *Wolbachia*-infected males, and recent work has demonstrated that WO prophage-associated loci cause CI (Beckmann and Fallon 2013; Beckmann *et al*. 2017; LePage *et al*. 2017; Beckmann *et al*. 2019). Although reproductive manipulations are common, some *Wolbachia* show little or no reproductive manipulation (*e.g.*, *w*Mel in *D. melanogaster*, Hoffmann 1988; Hoffmann *et al.* 1994; Kriesner *et al.* 2016; *w*Mau in *D. mauritiana*, Giordano *et al*. 1995; Meany *et al.* 2019; *w*Au in *D. simulans*, Hoffmann *et al.* 1996; *w*Suz in *D. suzukii* and *w*Spc in *D. subpulchrella*, Hamm *et al*. 2014; Cattel *et al*. 2018). 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; Kriesner and Hoffmann 2018). 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.* 2019), 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

3

61   cladogenic acquisition seems to be the rule for the obligate *Wolbachia* found in filarial

62   nematodes (Bandi *et al.* 1998); and there are also examples in at least two insect clades, *Nasonia*

63   wasps (Raychoudhury *et al.* 2009) and *Nomada* bees (Gerth and Bleidorn 2016). In contrast,

64   infections can be relatively young and acquired through introgressive or horizontal transfer

65   (Raychoudhury *et al.* 2009; Schuler *et al.* 2016; Conner *et al.* 2017; Turelli *et al.* 2018).

66   Comparisons of host nuclear and mitochondrial genomes with the associated *Wolbachia*

67   genomes enable discrimination among cladogenic, introgressive, and horizontal acquisition

68   (Raychoudhury *et al.* 2009; Turelli *et al.* 2018; see Figure S1). Concordant nuclear,

69   mitochondrial, and *Wolbachia* cladograms—including consistent divergence-time estimates for

70   all three genomes—support cladogenic acquisition and co-divergence of host and *Wolbachia*

71   lineages. Concordant *Wolbachia* and mitochondrial phylogenies and consistent *Wolbachia* and

72   mitochondrial divergence-time estimates that are more recent than nuclear divergence support

73   introgressive acquisition. In this case, mitochondrial and *Wolbachia* relationships may or may

74   not recapitulate the host phylogeny. Finally, if *Wolbachia* diverged more recently than either

75   nuclear or mitochondrial genomes, horizontal transfer (or paternal transmission; Hoffmann and

76   Turelli 1988; Turelli and Hoffmann 1995) is indicated. This is often associated with discordance

77   between host and *Wolbachia* phylogenies (O'Neill *et al.* 1992; Rousset and Solignac 1995;

78   Turelli *et al.* 2018).

79   Introgressive *Wolbachia* acquisition may be common in *Drosophila*. About half of all

80   closely related *Drosophila* species have overlapping geographical ranges and show pervasive

81   evidence of reinforcement (Coyne and Orr 1989, 1997; Yukelevich 2012; Nosil 2013), indicating

82   that hybridization must be common (Turelli *et al.* 2014). Several instances of sporadic

83   contemporary hybridization and interspecific gene flow have been documented in the genus

84   (Carson *et al.* 1989; Shoemaker *et al.* 1999; Jaenike *et al.* 2006; Kulathinal *et al.* 2009; Garrigan

85   *et al.* 2012; Brand *et al.* 2013; Matute and Ayroles 2014; Lohse *et al.* 2015); but only two stable

86   hybrid zones have been well described, and both involve *D. yakuba*-clade species. In west

87   Africa, *D. yakuba* hybridizes with endemic *D. santomea* on the island of São Tomé, and with *D.*

88   *teissieri* on the island of Bioko (Lachaise *et al.* 2000; Llopart *et al.* 2005; Comeault *et al.* 2016;

89   Cooper *et al.* 2018). The ranges of *D. yakuba* and *D. teissieri* overlap throughout continental

90   Africa, but contemporary hybridization has not been observed outside of Bioko (Cooper *et al.*

91   2018). Genomic analyses support both mitochondrial and nuclear introgression in the *D. yakuba*

92   clade (Lachaise *et al.* 2000; Bachtrog *et al.* 2006; Llopart *et al.* 2014; Turissini and Matute 2017;

93   Cooper *et al.* 2018), and the *Wolbachia* infecting all three species (*w*San, *w*Tei, and *w*Yak) are at

94   intermediate frequencies and identical with respect to commonly used typing loci (Lachaise *et al.*

95   2000; Charlat *et al.* 2004; Cooper *et al.* 2017).

96   Horizontal *Wolbachia* transmission has been repeatedly demonstrated since its initial

97   discovery by O'Neill *et al.* (1992) (*e.g.*, Baldo *et al.* 2008; Schuler *et al.* 2016), and it may be

98   common in some systems (*e.g.*, Huigens *et al.* 2000; Huigens *et al.* 2004; Ahmed *et al.* 2015; Li

99   *et al.* 2017). Phylogenomic analyses indicate recent horizontal *Wolbachia* transmission, on the

100 order of 5,000–27,000 years, among relatively distantly related *Drosophila* species of the *D.*

101 *melanogaster* species group (Turelli *et al.* 2018); but there is little evidence for non-maternal

102 transmission within *Drosophila* species (Richardson *et al.* 2012; Turelli *et al.* 2018). Horizontal

103 *Wolbachia* transfer could occur via a vector (Vavre *et al.* 2009; Ahmed *et al.* 2015) and/or via

104 shared food sources during development (Huigens *et al.* 2000; Li *et al.* 2017). Rare paternal

105 *Wolbachia* transmission has been documented in *D. simulans* (Hoffmann and Turelli 1988;

106 Turelli and Hoffmann 1995). However, the most common mode of *Wolbachia* acquisition by

107 *Drosophila* species (and most other hosts) remains unknown, and all three modes seem plausible

108 in the *D. yakuba* clade (Lachaise *et al.* 2000; Bachtrog *et al.* 2006; Cooper *et al.* 2017).

109   Distinguishing acquisition via introgression versus horizontal or paternal transmission

110 requires estimating phylogenies and relative divergence times of mtDNA and *Wolbachia*

111 (Raychoudhury *et al.* 2009; Conner *et al.* 2017; Turelli *et al.* 2018). To convert sequence

112 divergence to divergence-time estimates requires understanding relative rates of divergence for

113 mtDNA, nuclear genomes and *Wolbachia*. If each genome followed a constant-rate molecular

114 clock, taxa with cladogenic *Wolbachia* transmission, such as *Nasonia* wasps (Raychoudhury *et*

115 *al.* 2009) and *Nomada* bees (Gerth and Bleidorn 2016), would provide reference rates for

116 calibration. However, like Langley and Fitch (1974), Turelli *et al.* (2018) found significantly

117 varying relative rates of *Wolbachia* and mtDNA divergence. This variation, which we document

118 across *Drosophila*, confounds attempts to understand *Wolbachia* acquisition, as discussed below.

119   The discovery of weak CI in the *D. yakuba* clade (Cooper *et al.* 2017) motivates detailed

120 comparative analysis of loci associated with CI (CI factors or *cifs*) in WO prophage regions of

121 *Wolbachia* genomes. Beckmann and Fallon (2013) first associated wPip_0282 and wPip_0283

122 proteins in *w*Pip-infected *Culex pipiens* with *Wolbachia*-modified sperm. Later work confirmed

123   that these proteins induce toxicity and produce rescue when expressed/co-expressed in

124   *Saccharomyces cerevisiae* (Beckmann *et al.* 2017), lending support to a toxin-antidote model of

125   CI (Beckmann *et al.* 2019; but see, Shropshire *et al.* 2019). Homologs of these genes in *w*Mel

126   (WD0631 and WD0632) recapitulate CI when transgenically expressed in *D. melanogaster*

127   (LePage *et al.* 2017), and transgenic expression of WD0631 in *D. melanogaster* rescues CI

128   (Shropshire *et al.* 2018). A distantly related WO-prophage-associated pair, present in *w*Pip,

129   *w*Pip_0294 and *w*Pip_0295, causes similar toxicity/rescue in *S. cerevisiae* (Beckmann *et al.*

130   2017), and causes CI when placed transgenically into a *D. melanogaster* background

131   (unpublished data presented by Mark Hochstrasser at the 2018 *Wolbachia* meeting in Salem,

132   MA). We adopt Beckmann *et al.* (2019)'s nomenclature, which assigns names based on

133   enzymatic activity of the predicted toxin [deubiquitylase (DUB) and nuclease (Nuc)], with

134   superscripts denoting focal *Wolbachia* strains when needed. Specifically, we refer to *w*Pip_0282-

135   *w*Pip_0283 and the wPip_0294-wPip_0295 pairs as *cidA-cidB*^{wPip} and *cinA-cinB*^{wPip},

136   respectively; and we refer to WD0631-WD632 as *cidA-cidB*^{wMel}. This distinguishes *cid* (CI-

137   inducing DUB) from *cin* (CI-inducing Nuc) pairs, with the predicted antidote and toxin denoted

138   "A" and "B", respectively (Beckmann *et al.* 2019). We acknowledge ongoing disagreement in

139   the literature and direct readers to Beckmann *et al.* (2019) and Shropshire *et al.* (2019) for

140   details. However, none of these debates on terminology or mechanism affect our findings.

141   Here, we use host and *Wolbachia* genomes from the *D. yakuba* clade to demonstrate

142   introgressive and horizontal *Wolbachia* acquisition. General concordance of mitochondrial and

143   *Wolbachia* phylogenies indicates that horizontal acquisition is rare within this clade. However,

144   tests involving divergence-time estimates are complicated by varying relative rates and patterns

145   of *Wolbachia*, mtDNA, and nuclear sequence divergence, as illustrated by data from more

146   distantly related *Drosophila* (Clark *et al.* 2007). Finally, we demonstrate that *cid* loci underlying

147   CI in closely related *w*Mel (LePage *et al.* 2017) are disrupted in all *D. yakuba*-clade *Wolbachia*.

148   However, these *Wolbachia* also contain a set of *cin* loci, absent in *w*Mel, but very similar to

149   those found in *w*Pip (Beckmann *et al.* 2017), a B-group *Wolbachia* strain that diverged 6–46

150   million years ago (mya) from A-group *w*Yak and *w*Mel (Werren *et al.* 1995; Meany *et al.* 2019).

151   This is the first discovery of two sets of loci implicated in CI co-occuring within the same

152   prophage region. Several analyses implicate *Wolbachia*-specific insertion sequence (IS)

153   transposable elements, specifically ISWpi1, in the horizontal transfer of these loci (and

154 surrounding regions) between distantly related *Wolbachia*. Horizontal movement of

155 incompatibility factors between prophage regions of *Wolbachia* variants adds another layer to

156 what is already known about horizontal movement of prophages within and between *Wolbachia*

157 variants that themselves move horizontally between host species.

158

159 **MATERIALS AND METHODS**

160

161 **Genomic data**

162 The *D. yakuba*-clade isofemale lines included in our study were sampled over several years in

163 west Africa (Comeault *et al.* 2016; Turissini and Matute 2017; Cooper *et al.* 2017). Each line

164 within each species used in our analyses exhibits little nuclear introgression (< 1%) (Turissini

165 and Matute 2017), and no hybrids were included. Reads from *D. yakuba* ($N = 56$), *D. santomea*

166 ($N = 11$), and *D. teissieri* ($N = 13$) isofemale lines were obtained from the data archives of

167 Turissini and Matute (2017) and aligned to the *D. yakuba* nuclear and mitochondrial reference

168 genomes (Clark *et al.* 2007) with bwa 0.7.12 (Li and Durbin 2009), requiring alignment quality

169 scores of at least 50. Because many of the read archives were single end, all alignments were

170 completed using single-end mode for consistency.

171

172 **mtDNA:** Consensus mtDNA sequences for each of the 80 lines were extracted with samtools v.

173 1.3.1 and bcftools v 1.3.1 (Li 2011). Coding sequences for the 13 protein-coding genes were

174 extracted, based on their positions in the *D. yakuba* reference. We also extracted the 13 protein-

175 coding genes from additional unique *D. yakuba* ($N = 28$) and *D. santomea* ($N = 15$)

176 mitochondrial genomes (Llopart *et al.* 2014), and from the *D. melanogaster* reference (Hoskins

177 *et al.* 2015). Genes were aligned using MAFFT v. 7 and concatenated (Katoh and Standley

178 2013). Lines identical across all 13 genes were represented by one sequence for the phylogenetic

179 analyses.

180

181 ***Wolbachia*:** Reads from each of the 80 Turissini and Matute (2017) lines were aligned using bwa

182 0.7.12 (Li and Durban 2009) to the *D. yakuba* reference genome (Clark *et al.* 2007) combined

183 with the *w*Mel reference genome (Wu *et al.* 2004). We calculated the average depth of coverage

184 across the *w*Mel genome. We considered lines with < 1× coverage uninfected and lines with

185   >10× coverage infected. No lines had between 1× and 10× coverage. To test our genomic

186   analyses of infection status, we used a polymerase chain reaction (PCR) assay on a subset of

187   lines. We extracted DNA using a "squish" buffer protocol (Gloor *et al.* 1993), and infection

188   status was determined using primers for the *Wolbachia*-specific *wsp* gene (Braig *et al.* 1998;

189   Baldo *et al.* 2006). We amplified a *Drosophila*-specific region of chromosome *2L* as a positive

190   control (Kern *et al.* 2015) (primers are listed in Supplemental Material, Table S1). For each run,

191   we also included a known *Wolbachia*-positive line and a water blank as controls.

192   To produce a draft *Wolbachia* "pseudoreference" genome for each species, we first

193   determined the isofemale line from each species that produced the greatest average coverage

194   depth over *w*Mel. We trimmed the reads with Sickle v. 1.33 (Joshi and Fass 2011) and assembled

195   with ABySS v. 2.0.2 (Jackman *et al.* 2017). *K* values of 51, 61…91 were tried. Scaffolds with

196   best nucleotide BLAST matches to known *Wolbachia* sequences with E-values less than 10^{-10}

197   were extracted as the draft *Wolbachia* assembly. For each species, the assembly with the highest

198   N50 and fewest scaffolds was kept as our *Wolbachia* pseudoreference genome for that host

199   species, denoted *w*Yak, *w*San, and *w*Tei (Table S2). To assess the quality of these three draft

200   assemblies, we used BUSCO v. 3.0.0 to search for homologs of the near-universal, single-copy

201   genes in the BUSCO proteobacteria database (Simão *et al.* 2015). For comparison, we followed

202   Conner *et al.* (2017) and performed the same search on the reference genomes for *w*Mel (Wu *et*

203   *al.* 2004), *w*Ri (Klasson *et al.* 2009), *w*Au (Sutton *et al.* 2014), *w*Ha and *w*No (Ellegaard *et al.*

204   2013) (Table S3).

205   Using these draft *Wolbachia* pseudoreference genomes, reads from all other genotypes were

206   aligned to the *D. yakuba* reference (nuclear and mitochondrial) plus the species-specific

207   *Wolbachia* draft assembly with bwa 0.7.12 (Li and Durbin 2009). Consensus sequences were

208   extracted with samtools v. 1.3.1 and bcftools v. 1.3.1 (Li 2011).

209   To test whether the choice of *Wolbachia* pseudoreference influenced the *Wolbachia* draft

210   sequence obtained for each infected line, we arbitrarily selected three infected lines of each host

211   species and aligned reads from each of those lines independently to our *w*Yak, *w*San, and *w*Tei

212   pseudoreferences. This resulted in 27 alignments (9 per host species). Among the three

213   *Wolbachia* consensus sequences generated for each line, we assessed single-nucleotide variants

214   between different mappings of the same line within loci used for downstream analyses.

215   Irrespective of the pseudoreference used, we obtained identical *Wolbachia* sequences. This

8

216   indicates the robustness of our approach to generating population samples of draft *Wolbachia*

217   genomes, without having a high-quality *Wolbachia* reference genome from any of the host

218   species. Because Llopart *et al.* (2014) released only assembled mitochondrial genomes, we could

219   not assess the *Wolbachia* infection status of their lines.

220

221   **Loci for phylogenetic and comparative-rate analyses**

222   ***Wolbachia* genes:** For each phylogenetic analysis of *Wolbachia* data, the draft genomes were

223   annotated with Prokka v. 1.11 (Seemann 2014), which identifies homologs to known bacterial

224   genes. To avoid pseudogenes and paralogs, we used only genes present in a single copy and with

225   identical lengths in all of the sequences analyzed. Genes were identified as single copy if they

226   uniquely matched a bacterial reference gene identified by Prokka v. 1.11. By requiring all

227   homologs to have identical length in all of our draft *Wolbachia* genomes, we removed all loci

228   with indels across any of our sequences.

229   Given that many loci accumulate indels over time, the number of loci included in our

230   phylogenetic analyses depended on the number of strains included. We first estimated

231   phylograms for the A-group *Wolbachia* from: the *D. yakuba*-clade, *w*Mel from *D. melanogaster*

232   (Wu *et al.* 2004), *w*Inc from *Drosophila incompta* (Wallau *et al.* 2016), *w*Suz from *D. suzukii*

233   (Siozios et al. 2013), *w*Ana from *D. anannasae* (Choi *et al.* 2015), *Wolbachia* that infect *Nomada*

234   bees (*w*NFe, *w*NPa, *w*NLeu, and *w*NFa; Gerth and Bleidorn 2016), and *Wolbachia* that infect *D.*

235   *simulans* (*w*Ri, *w*Au and *w*Ha; Klasson *et al.* 2009; Sutton *et al.* 2014; Ellegaard *et al.* 2013). We

236   also analyzed distantly related B-group *Wolbachia*: *w*No from *D. simulans* (Ellegaard *et al.*

237   2013), *w*Pip_Pel from *Culex pipiens* (Klasson *et al.* 2008), and *w*AlbB from *Aedes albopictus*

238   (Mavingui *et al.* 2012). To increase our phylogenetic resolution within the *D. yakuba Wolbachia*

239   clade (by increasing the number of loci), we estimated a phylogram that included only *D.*

240   *yakuba*-clade *Wolbachia* and *w*Mel. Our phylogram with both A- and B-group *Wolbachia*

241   included 146 genes (containing 115,686 bp), and our phylogram that included only *D. yakuba*-

242   clade *Wolbachia* and *w*Mel included 643 genes (containing 644,586 bp).

243   We also constructed an absolute chronogram to estimate divergence between *D. yakuba*-

244   clade *Wolbachia* strains. To illustrate how divergence-time estimates vary with changing patterns

245   of molecular divergence, we estimated two additional *Wolbachia* chronograms: one that

246   considered only the *w*Mel reference plus the two most-diverged *w*Mel lines included in

247 Richardson *et al.* (2012), and one that included these three *w*Mel variants and our *D. yakuba*-

248 clade *Wolbachia*. Our chronogram that included only *D. yakuba*-clade *Wolbachia* included 678

249 genes (containing 695,118 bp), our chronogram that included only *w*Mel *Wolbachia* included

250 692 genes (containing 709,599 bp), and our chronograms with *w*Mel plus *D. yakuba*-clade

251 *Wolbachia* included 621 genes (containing 624,438 bp). Independent estimates were obtained

252 using relaxed-clock analyses with $\Gamma(2,2)$ and $\Gamma(7,7)$ branch-rate priors. We also estimated

253 divergence between *D. yakuba*-clade *Wolbachia* and *w*Mel using a strict-clock analysis,

254 corresponding to $\Gamma(n,n)$ as $n \rightarrow \infty$.

255

256 **mtDNA and nuclear genes from diverse *Drosophila*:** To assess variation in the relative rates of

257 divergence of mitochondrial and nuclear genes, we analyzed the canonical 12 *Drosophila*

258 genomes (Clark *et al.* 2007) plus *D. suzukii* (Chiu *et al.* 2013). We excluded *D. sechellia* and *D.*

259 *persimilis* which show evidence of introgression with *D. simulans* and *D. pseudoobscura*,

260 respectively (Kulathinal *et al.* 2009; Schrider *et al.* 2018; Brand *et al.* 2013). Coding sequences

261 for 20 nuclear genes used in the analyses of Turelli *et al.* (2018) (*aconitase, aldolase, bicoid,*

262 *ebony, enolase, esc, g6pdh, glyp, glys, ninaE, pepck, pgi, pgm, pic, ptc, tpi, transaldolase, white,*

263 *wingless*, and *yellow*) were obtained from FlyBase for each species. The genes were aligned with

264 MAFFT v. 7 (Katoh and Standley 2013). Coding sequences for the 13 protein-coding

265 mitochondrial genes in the inbred reference strains were also obtained from FlyBase for each

266 species and were aligned with MAFFT v. 7 and concatenated.

267

268 **Phylogenetic analyses**

269 All of our analyses used RevBayes v. 1.0.9 (Hohna *et al.* 2016), following the procedures of

270 Turelli *et al.* (2018). For completeness, we summarize those methods below. For additional

271 details on the priors and their justifications, consult Turelli *et al.* (2018). Four independent runs

272 were performed for each phylogenetic tree we estimated; and in all cases, the runs converged to

273 the same topologies. Nodes with posterior probability less than 0.95 were collapsed into

274 polytomies.

275

276 ***Wolbachia* phylograms:** We estimated a phylogram for A- and B-group *Wolbachia* and for only

277 *D. yakuba*-clade and *w*Mel *Wolbachia* using the same methodology as Turelli *et al.* (2018). We

10

278   used a GTR + Γ model with four rate categories, partitioning by codon position. Each partition

279   had an independent rate multiplier with prior Γ(1,1) (i.e., Exp(1)), as well as stationary

280   frequencies and exchangeability rates drawn from flat, symmetrical Dirichlet distributions (i.e.,

281   Dirichlet(1,1,1...). The model used a uniform prior over all possible topologies. Branch lengths

282   were drawn from a flat, symmetrical Dirichlet distribution, thus they summed to 1. Since the

283   expected number of substitutions along a branch equals the branch length times the rate

284   multiplier, the expected number of substitutions across the entire tree for a partition is equal to

285   the partition's rate multiplier.

286

287   ***Wolbachia* chronograms:** We first created a relaxed-clock relative chronogram with the root

288   age fixed to 1 using the GTR + Γ model, partitioned by codon position, using the same birth-

289   death prior as Turelli *et al.* (2018). Each partition had an independent rate multiplier with prior

290   Gamma(1,1), as well as stationary frequencies and exchangeability rates drawn from flat,

291   symmetrical Dirichlet distributions. The branch-rate prior for each branch was

292   Γ(2,2), normalized to a mean of 1 across all branches (Table S4). We also tried a strict-clock tree

293   and a relaxed-clock tree with branch-rate prior Γ(7,7), which produced no significant differences.

294   We used the scaled distribution $\Gamma(7,7) \times 6.87 \times 10^{-9}$ to model substitutions per third-position site

295   per year. This transforms the relative chronogram into an absolute chronogram. This scaled

296   distribution was chosen to replicate the upper and lower credible intervals of the posterior

297   distribution estimated by Richardson *et al*. (2012), assuming 10 *Drosophila* generations per year,

298   normalized by their median substitution-rate estimate. Branch lengths in absolute time were

299   calculated as the relative-branch length times the third-position rate multiplier divided by the

300   substitutions per third-position site per year estimate above.

301   We illustrate how divergence-time estimates depend on changing patterns of *Wolbachia*

302   molecular evolution observed over different time scales, specifically the relative rates of third-

303   site substitutions versus first- and second-site substitutions. We compare divergence times

304   estimated for variants within *melanogaster*-group host species, with a time scale of only

305   hundreds or thousands of years, with the same divergence times estimated when more distantly

306   related *Wolbachia*, with divergence times over tens of thousands of years, are included in the

307   analyses. We present three separate analyses: one using only *D. yakuba*-clade *Wolbachia*

308   variants, a second using only *w*Mel variants analyzed by Richardson *et al.* (2012), and a third

309   simultaneously analyzing both sets of data.

310

311   ***Drosophila* phylogeny for 11 references species:** We estimated a phylogram from our 20

312   nuclear loci using the GTR + Γ model, with partitioning by gene and codon position. The model

313   was identical to the *Wolbachia* phylogram model above except for the partitioning (there are too

314   few *Wolbachia* substitutions to justify partitioning by gene).

315

316   **mtDNA phylogeny for the 11 reference species**: We estimated a phylogram from the 13

317   protein-coding mitochondrial loci using the GTR + Γ model, with partitioning only by codon

318   position. The model was identical to the *Wolbachia* phylogram model above.

319

320   **mtDNA chronogram for the *D. yakuba* clade:** We estimated a relative chronogram for the *D.*

321   *yakuba*-clade mitochondria from the 13 protein-coding loci using the GTR + Γ model,

322   partitioning only by codon position. To test the sensitivity of our results to priors, we ran a strict

323   clock, a relaxed clock with a Γ(7,7) branch-rate prior, and a relaxed clock with a Γ(2,2) branch-

324   rate prior, corresponding to increasing levels of substitution-rate variation across branches. The

325   model was identical to the *Wolbachia* chronogram model above, except that we did not

326   transform it into an absolute chronogram.

327

328   **Ratios of divergence rates for mtDNA versus nuclear loci:** To quantify ratios of mtDNA to

329   nuclear substitution rates, we estimated relative substitution rates for host nuclear genes versus

330   mtDNA using the GTR + Γ model. The (unrooted) topology was fixed to the consensus topology

331   for the nuclear and mitochondrial data. The data from 20 nuclear loci were partitioned by locus

332   and codon position, the mtDNA data were partitioned only by codon position. All partitions

333   shared the same topology, but the nuclear partitions were allowed to have branch lengths

334   different from the mtDNA partitions. The sum of the branch lengths for each partition was scaled

335   to 1. Assuming concurrent nuclear-mtDNA divergence (because we used only species showing

336   no evidence of introgression), we imposed the same absolute ages for all nodes of the nuclear

337   and mtDNA chronograms. For each nuclear locus, the third-position rate-ratio for the mtDNA

338   versus nuclear genomes was calculated as: (mitochondrial branch length × mitochondrial 3rd

12

339 position rate multiplier)/(nuclear branch length × nuclear 3rd position rate multiplier). We

340 summarized the relative rates of mtDNA versus nuclear substitutions along each branch using

341 the arithmetic average of the 20 ratios obtained from the individual nuclear loci.

342

343 **Introgressive versus horizontal *Wolbachia* transfer–concordance of phylograms:** Horizontal

344 transfer of *Wolbachia* was initially detected by discordance between *Wolbachia* and host

345 phylogenies (O'Neill *et al*. 1992). When considering samples within species and closely related

346 species, a comparison of mitochondrial and *Wolbachia* phylogenies provides a natural test for

347 horizontal transmission. We first look for significant discrepancies between strongly supported

348 nodes in the mitochondrial versus *Wolbachia* phylogenies. To test for less obvious differences,

349 we follow Richardson *et al*. (2012) and compute Bayes Factors to assess the support for models

350 that assume that mitochondrial versus *Wolbachia* follow the same topology versus distinct

351 phylogenies. For these calculations, we omit san-Quija37 due to its clear discordance.

352   We calculated the marginal likelihood in RevBayes for two models; one with a shared

353 mitochondrial and *Wolbachia* phylogeny, another with independent topologies. The mtDNA and

354 *Wolbachia* were each partitioned by codon position, for a total of six partitions. In the shared

355 phylogeny model, all six share the same topology; in the independent model, mitochondria and

356 *Wolbachia* have separate topologies. All priors were the same as the *Wolbachia* phylogram

357 model. Lines that were identical across the mtDNA and *Wolbachia* were collapsed into one

358 sample. We ran two independent replicates of each model with 50 stepping stones per run. The

359 Bayes factor is computed as the difference between the marginal likelihoods of each model.

360

361 **Introgressive versus horizontal *Wolbachia* transfer–relative rates:** Within species, the

362 phylogenies of mitochondria and *Wolbachia* are fully resolved. Thus we consider an alternative

363 approach for distinguishing between introgressive versus horizontal transfer that depends on

364 estimating divergence times for *Wolbachia* genomes and host mtDNA. This is complicated by

365 variation in the relative rates of mtDNA versus *Wolbachia* divergence and by systematic changes

366 over time in the relative rates of substitutions at the three codon positions for *Wolbachia*.

367 Following the procedures in Turelli *et al.* (2018), we estimated the mtDNA-to-*Wolbachia* third-

368 position substitution-rate ratio for each branch in the *D. yakuba* clade. For each analysis, we

369 computed the marginal likelihood of the model where all branches shared the same ratio and the

13

370   same model except allowing different ratios on each branch. We then calculated Bayes factors

371   (i.e. differences in the log of the marginal likelihoods) to determine which model was favored.

372   We repeated these analyses after including three *w*Mel-infected *D. melanogaster* lines with the

373   *D. yakuba*-clade dataset. Because *D. yakuba*-clade species and *D. melanogaster* do not produce

374   fertile hybrids (Sánchez and Santamaria 1997; Turissini *et al.* 2017), introgressive transfer of

375   *Wolbachia* is not possible between these species. Hence, including *w*Mel provides a control for

376   our proposed test of introgressive transfer. For these controls, we calculated the mitochondrial

377   versus *Wolbachia* substitution-rate ratio for the interspecific branches separately, regardless of

378   whether the shared rate-ratio model was favored.

379

380   **_Wolbachia_ loci associated with CI**

381   *Wolbachia* that infect all three *D. yakuba*-clade hosts cause weak intra-and interspecific CI

382   (Cooper *et al.* 2017), but the genetic basis of CI in this clade remains unknown. We used

383   tBLASTN to search for *cif* homologs in each of our *D. yakuba*-clade *Wolbachia* assemblies,

384   querying *cidA-cidB* variants and the *cinA-cinB*$^{w\text{Ri}}$ pair (wRi_006720 and wRi_006710) found in

385   *w*Ri and some *w*Ri-like *Wolbachia* (Beckmann and Fallon 2013; Turelli *et al.* 2018), and the

386   *cinA-cinB*$^{w\text{Pip}}$ pair (Beckmann and Fallon 2013; Beckmann *et al.* 2017; LePage *et al.* 2017;

387   Lindsey *et al.* 2018). We identified both *cid* and *cin* homologs, but we found no close matches to

388   the *cinA-cinB*$^{w\text{Ri}}$ pair (denoted Type II *cif* loci by Lepage *et al.* 2017) in any of our genomes. For

389   all of the samples, we extracted consensus sequences from our assemblies and alignments for the

390   *cidA-cidB*$^{w\text{Yak-clade}}$ and *cinA-cinB*$^{w\text{Yak-clade}}$ gene pairs. The genes were aligned with MAFFT v. 7

391   (Katoh and Standley 2013). We examined variation in these regions relative to *cidA-cidB*$^{w\text{Mel}}$ and

392   *cinA-cinB*$^{w\text{Pip}}$ respectively.

393   Because we unexpectedly identified homologs of *cinA-cinB* in all *w*Yak-clade *Wolbachia*,

394   we took additional measures to understand the origin and placement of these loci in the genomes.

395   The *cinA-cinB*$^{w\text{Yak-clade}}$ open reading frames are located on a ~11,500 bp scaffold in the

396   fragmented *w*Yak assembly (*w*Yak scaffold "702380" from ABySS output—we use quotes to

397   indicate names assigned by ABySS). The first ~4,000 bp of this scaffold contain *cinA* and *cinB*

398   genes that have ~97% identity with those in *w*Pip. We performed a BLAST search using all

399   contigs in the *w*Yak assembly as queries against the *w*Mel genome (Camacho *et al.* 2009). The

400   *cinA-cinB*$^{w\text{Yak-clade}}$ scaffold was placed on the *w*Mel genome (~617,000–623,000) adjacent to and

14

401   downstream of the wYak contig containing $cidA$-$cidB^{w\text{Yak-clade}}$ loci (wYak contig "187383").

402   However, only ~7,000 bp of the 11,500 bp of the scaffold containing $cinA$-$cinB^{w\text{Yak-clade}}$ align to

403   this region of wMel. The ~4,000 bp sequence that contains the ORFs for $cinA$-$cinB^{w\text{Yak-clade}}$ has

404   no significant hit against the wMel genome.

405   To verify the placement of the scaffold containing $cinA$-$cinB^{w\text{Yak-clade}}$, we first performed

406   targeted assembly using the $cinA$-$cinB^{w\text{Yak-clade}}$ contig and the adjacent contigs as references

407   (Hunter *et al*. 2015). Iterative targeted assembly often extends assembled scaffolds (Langmead

408   and Salzberg 2012). Reads are mapped to the target of interest, and the mapped reads are then

409   assembled using SPADES (Bankevich *et al*. 2012; Nurk *et al*. 2013). The newly assembled

410   scaffolds serve as the target in a subsequent round of mapping. The procedure is repeated until

411   no new reads are recruited to the assembled scaffold from the previous round. The extended

412   scaffolds enabled us to merge flanking wYak contigs downstream of the $cinA$-$cinB^{w\text{Yak-clade}}$

413   scaffold but failed to connect the fragment with the $cidA$-$cidB^{w\text{Yak-clade}}$ contig. We used PCR to

414   amplify the intervening region (Supplemental Information, Table S1), and Sanger sequencing to

415   evaluate this product (Sanger *et al.* 1977).

416   Subsequent mapping of paired-end reads to the merged scaffolds confirmed the correct

417   order and orientation of the contigs containing the $cidA$-$cidB^{w\text{Yak-clade}}$ and $cinA$-$cinB^{w\text{Yak-clade}}$

418   genes. Observed, uncorrected, pairwise distances between focal regions in wYak, wMel, and

419   wPip and between focal regions in wYak, wPip, wAlbB, and wNleu were calculated using a

420   sliding window (window size = 200 bp, step size = 25 bp).

421

Data Availability

423   The *Wolbachia* assemblies (BioProject PRJNA543889, Accession VCEF00000000) and Sanger

424   sequences (MK950151 and MK950152) are archived in GenBank. All three assemblies were

425   manually corrected using Sanger sequence to accurately represent variation in the *cin* region. All

426   relevant code will be uploaded to Dryad prior to publication.

427

428

429

430

431

432 **RESULTS**

433

434 **Unidirectional introgression of *D. yakuba* mitochondria into *D. santomea***

435 Previous analyses suggest that *D. yakuba* and *D. santomea* carry very similar mtDNA due to

436 ongoing hybridization and introgression (Lachaise *et al.* 2000; Bachtrog *et al.* 2006; Llopart *et*

437 *al.* 2014). However, our mtDNA relative chronogram (Figure 1), based on mitochondrial whole-

438 proteome data sampled from throughout the ranges of all three species, supports three mtDNA

439 clades that largely agree with the nuclear topology—*D. teissieri* mtDNA sequences are outgroup

440 to sister *D. yakuba* and *D. santomea* mtDNA (Figure 1). Consistent with introgression, 12 out of

441 26 *D. santomea* isofemale lines have *D. yakuba*-like mtDNA (indicated by blue branches and

442 letters in Figure 1). Yet, our samples showed no evidence of *D. santomea* mtDNA introgressed

443 into *D. yakuba* and no mtDNA introgression involving *D. teissieri*. The 12 *D. santomea*

444 isofemales with introgressed *D. yakuba*-like mtDNA were sampled from both within ($N = 8$) and

445 outside ($N = 4$) the well-described Pico de São Tomé hybrid zone (Llopart et al. 2014; Turissini

446 and Matute 2017). The *D. santomea* hybrid-zone samples are indicated by "HZ" in Figure 1 in

447 blue, and they are found throughout the clade that includes all *D. yakuba* mtDNA. This suggests

448 that hybridization occurs in other areas of the island or that introgressed *D. santomea* genotypes

449 migrate outside of the hybrid zone.

450   Previous analysis of nuclear genomes (Turissini and Matute 2017) also found more

451 introgression from *D. yakuba* into *D. santomea* when populations of *D. yakuba* from near the

452 Gulf of Guinea, Cameroon, and Kenya were included; however, excluding Cameroon and Kenya

453 indicated similar amounts of introgression in each direction. Matings between *D.*

454 *santomea* females and *D. yakuba* males are rare relative to the reciprocal cross (Coyne *et al.*

455 2002, Matute 2010). When these matings do occur, F1 females produce fewer progeny than do

456 F1 females produced by *D. yakuba* females and *D. santomea* males (Matute and Coyne 2010).

457 F1 hybrids produced by *D. santomea* females also have a shortened lifespan (Matute and Coyne

458 2010), due to copulatory wounds inflicted by *D. yakuba* males during mating (Kamimura 2012).

459 These observations are consistent with our finding of preferential introgression of *D.*

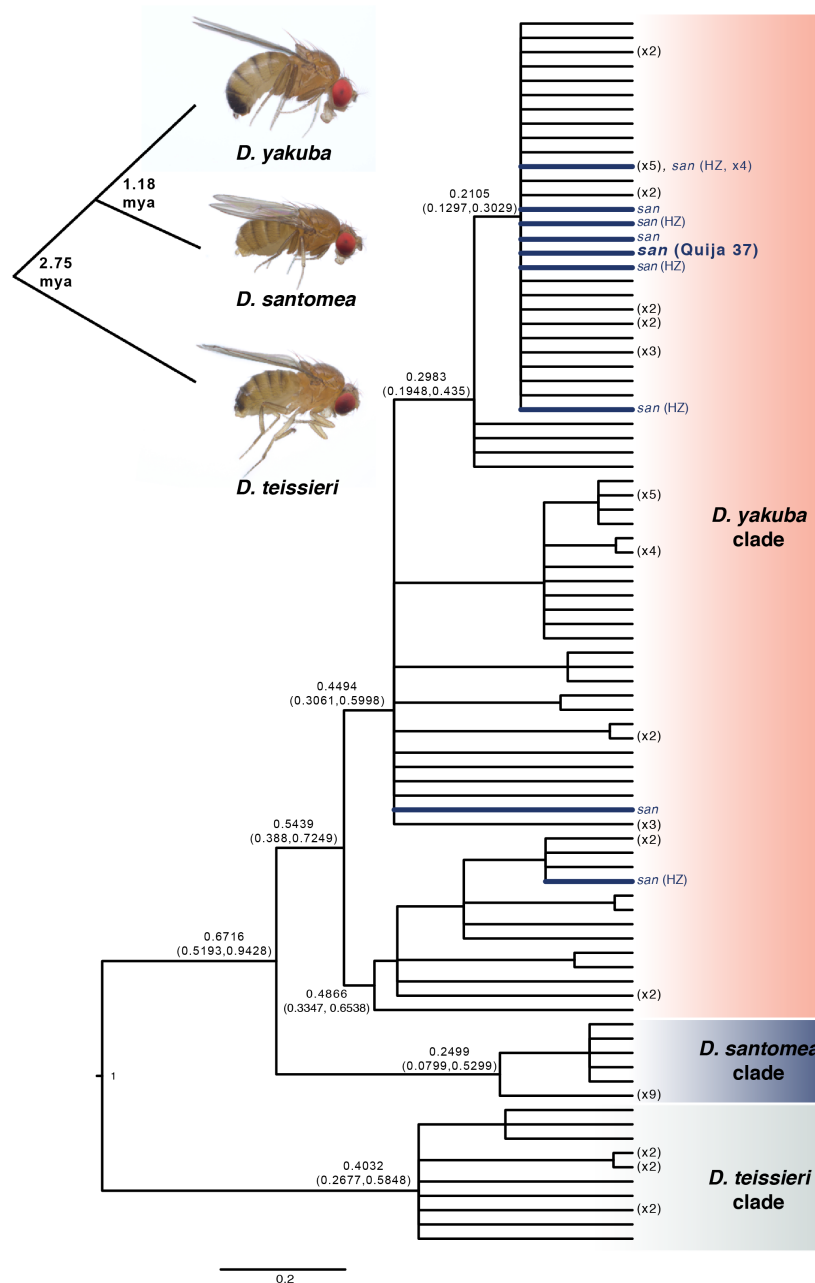460 *yakuba* mitochondria into *D. santomea* backgrounds.

461

462

16

463

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. (The 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.

17

Wolbachia **frequencies and draft genomes**

464 As expected, we found *Wolbachia* in all three *D. yakuba*-clade species sampled by Turissini and

465 Matute (2017). For *D. yakuba,* 21 of 56 lines were infected, yielding an infection frequency of *p*

466 = 0.36, with 95% binomial confidence interval (0.25, 0.51). For *D. santomea*, 10 of 11 were

467 infected, *p* = 0.91 (0.59, 1.0); and for *D. teissieri*, 11 of 13 were infected, *p* = 0.85 (0.55, 0.98).

468 Additional frequency estimates are reported in Cooper *et al.* (2017), which found that *Wolbachia*

469 frequencies vary through time and space in west Africa. All PCR tests for *Wolbachia* matched

470 our coverage-based genomic analyses of infection status.

471 The average coverage across the genome, calculated as the total number of bases aligned to

472 the *w*Mel reference divided by its length, was 1,940 for our *w*Yak pseudoreference genome (yak-

473 CY17C), 40 for *w*San (san-Quija630.39), and 489 for *w*Tei (teis-cascade_4_2). These

474 pseudoreference genomes were included in the phylogram that includes A- and B-group

475 *Wolbachia* (Figure 2), and they cluster with the *Wolbachia* from their respective hosts (Figure 3),

476 as expected if introgression is rare. In general, the *w*San data are lower quality due to the

477 relatively small size of the *D. santomea* libraries (Turissini and Matute 2017). The scaffold

478 count, N50, and total assembly sizes are reported in Table S2.

Introgressive and horizontal *Wolbachia* transfer among hosts

481 Our phylogram that includes A- and B-group *Wolbachia* places the three weak-CI-causing *D.*

482 *yakuba*-clade *Wolbachia* together and sister to *w*Mel in the A group (Figure 2A). In contrast, *D.*

483 *simulans* carries diverse *Wolbachia* that do (*w*Ri, *w*Ha, *w*No) and do not (*w*Au) cause CI,

484 spanning *Wolbachia* groups A (*w*Ri, *w*Ha, *w*Au) and B (*w*No). All *w*San, *w*Yak, and *w*Tei

485 *Wolbachia* variants included in our analysis are identical at the MLST loci often used to type

486 *Wolbachia* variants (Cooper *et al.* 2017; Baldo *et al.* 2006). Due to the relatively small number

487 of genes that meet our inclusion criteria (146 genes across 115,686 bp), the resulting phylogram

488 does not resolve relationships among *w*San, *w*Yak, and *w*Tei (Figure 2A). Our phylogram that

489 includes only *D. yakuba*-clade and *w*Mel *Wolbachia* (based on 643 genes across 644,586 bp)

490 resolves these relationships, placing *w*Tei variants outgroup to sister *w*San and *w*Yak in distinct

491 clades (Figure 2B). We observe 0.0039% third-position pairwise differences between *w*Tei and

492 *w*Yak and between *w*Tei and *w*San, and 0.0017% between *w*San and *w*Yak. The third-position

493 pairwise differences between any of the *D. yakuba*-clade *Wolbachia* and *w*Mel is 0.11%. For

18

495 reference, the third-position pairwise difference between *w*Yak and *w*Ri is 2.9%, whereas *w*Ri

496 and *w*Suz differ by only 0.014% (Turelli *et al.* 2018).
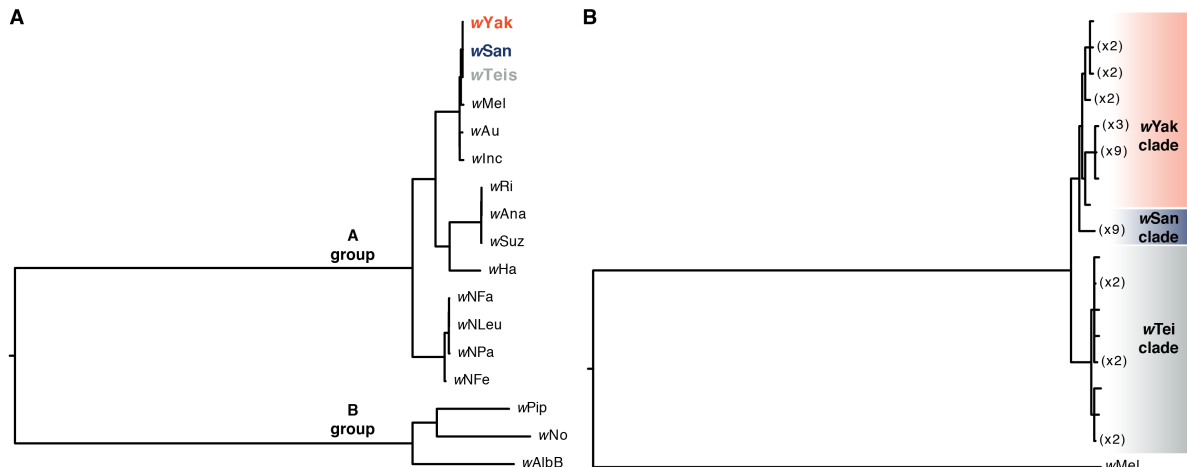
497

Figure 2 A) Bayesian phylogram placing the *D. yakuba*-clade *Wolbachia* in the context of other A-, and also B-group, *Wolbachia*. *w*Pip, *w*No, and *w*AlbB 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 bp. 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.

498 Our absolute chronogram that includes *w*Mel and *D. yakuba*-clade *Wolbachia* also indicates

499 three monophyletic groups (Figure 3), whose topology—(*w*Tei, (*w*Yak, *w*San))—agrees with the

500 nuclear topology of the hosts (Turissini and Matute 2017). However, one *D. santomea* line, *D.*

501 *santomea* Quija 37, sampled from the south side of São Tomé (Matute 2010), carries *Wolbachia*

502 identical to the other eight *w*San samples across 695,118 bp, yet its mtDNA is nested in the *D.*

503 *yakuba* mtDNA cluster (bold blue in Figure 1). This single example of *Wolbachia*-mtDNA

504 phylogenetic discordance can be explained by either horizontal *Wolbachia* transfer or paternal

505 transfer of *Wolbachia* or mtDNA. While rare, paternal *Wolbachia* transmission has been

506 observed in other *Drosophila* (Hoffmann and Turelli 1988; Turelli and Hoffmann 1995), as has

507 paternal transmission of mtDNA (Kondo *et al.* 1990). The only other *Wolbachia*-screened *D.*

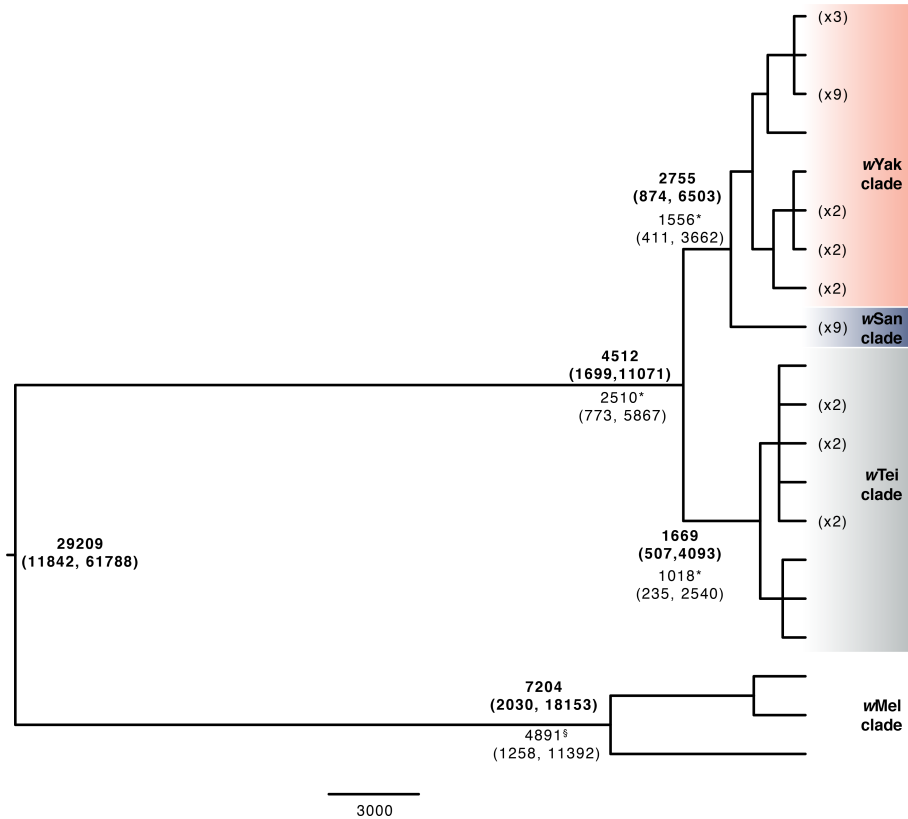508 *santomea* isofemale line with introgressed mtDNA was *Wolbachia* uninfected.

19

509

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, although credible intervals overlap in each case. 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).

510

511   In Figure 3, we present alternative estimates of the divergence times for *Wolbachia* within

512   and between *D. yakuba*-clade species, within *D. melanogaster*, and between the *D. yakuba* clade

513   and *D. melanogaster*. The estimates from the chronogram that includes all of the indicated

514   *Wolbachia* variants are in bold. To demonstrate how these time estimates depend on the

515   sequences included in the analyses (through estimates of relative divergence for each codon

20

516   position), estimates that included only *D. yakuba*-clade *Wolbachia* or only *w*Mel variants are

517   superimposed and not in bold. These estimates were all obtained using a relaxed-clock analysis

518   with a $\Gamma(2,2)$ branch-rate prior (results using $\Gamma(7,7)$ and a strict clock are mentioned below).

519   When considering only the *D. yakuba*-clade *Wolbachia* (Figure 3), we estimate the root at 2,510

520   years (95% credible interval = 773 to 5,867 years) and the *w*Yak-*w*San split at 1,556 years (95%

521   credible interval = 411 to 3,662 years). The *w*Mel chronogram estimates the most recent

522   common ancestor (MRCA) of *w*Mel at 4,890 years ago (95% credible interval = 1,258 to 11,392

523   years). Our *w*Mel result is consistent with the Richardson *et al.* (2012) estimate of 8,008 years

524   (95% credible interval = 3,263, 13,998). The alternative branch-rate $\Gamma(7,7)$ prior has little effect

525   on these divergence-times estimates, with credible intervals that overlap with those produced

526   using a $\Gamma(2,2)$ branch-rate prior (Figure 3 and Table S5). In contrast, Figure 3 shows that

527   including more divergent variants influences our time estimates, although the credible intervals

528   are overlapping in each case. Table S4 shows how estimates of the relative amounts of the

529   divergence for the three codon positions vary across the data sets. The underlying model for

530   divergence times assumes constant relative rates at the three positions. Hence, changes in relative

531   rates affect divergence-time estimates.

532   As noted above, our estimates of the time of the MRCA for *Wolbachia* within the *D. yakuba*

533   clade or within *D. melanogaster* are relatively insensitive to whether the branch-rate prior is

534    $\Gamma(2,2)$ or $\Gamma(7,7)$. The choice of genes used to estimate divergence also has little effect—maximal

535   distance between *D. yakuba*-clade strains was identical for analyses that included only the *D.*

536   *yakuba*-clade (*N* = 678 genes) or *w*Yak-clade plus *w*Mel (*N* = 621 genes) gene sets (*w*Yak-*w*Tei

537   distance = 0.0035%). The estimated divergence time between *w*Mel and the *D. yakuba*-clade

538   *Wolbachia* is less robust, although again the credible intervals generated using alternative

539   branch-rate priors are overlapping (Figure 3 and Table S5). A strict-clock analysis produces an

540   estimate of 72,612, with 95% credible interval (23,412, 132,276), again overlapping with our

541   relaxed-clock estimates. An alternative point estimate can be obtained from pairwise differences.

542   Averaging over the *D. yakuba*-clade sequences, we observe an average third-position difference

543   of 0.107% between the *D. yakuba*-clade *Wolbachia* and *w*Mel, or 0.0535% from tip to root. The

544   "short-term evolutionary rate" of divergence within *w*Mel estimated by Richardson *et al.* (2012)

545   produces a point estimate of 78,000 years, which overlaps with the credible interval of our

546   relaxed-tree estimate with a $\Gamma(7,7)$ branch-rate prior (Table S5). This molecular-clock estimate

21

547   closely agrees with our strict-clock chronogram analysis (Figure S2). Given that these estimates

548   are much shorter than the divergence time between their reproductively isolated hosts, horizontal

549   *Wolbachia* transmission must have occurred. As discussed below, the horizontal transmission

550   probably involved intermediate hosts.

551

552   **mtDNA and nuclear concordance across *Drosophila*:** The topologies of the nuclear and

553   mtDNA trees for the 11 reference *Drosophila* species are completely concordant (Figure 4) and

554   agree with the neighbor-joining result presented in Clark *et al*. (2007). We find the average ratio

555   of mtDNA to nuclear substitution rates across branches vary from about $1\times$ to $10\times$, consistent

556   with the results of Havird and Sloan (2016) across Diptera.

557

558   **mtDNA and *Wolbachia* concordance in the *D. yakuba* clade:** The discordance between the

559   placement of the mitochondria of *D. santomea* line Quija 37, which has mitochondria belonging

560   to the clade associated *D. yakuba* and its *Wolbachia*, which is clearly *w*San, demonstrates

561   horizontal or paternal transmission. Apart from this, we find no nodes in either the mitochondrial

562   or *Wolbachia* tree with posterior probability at least 0.95 that is inconsistent with a strongly

563   supported node (i.e., posterior probability at least 0.95) in the other tree (Figure S3). For a more

564   refined test of concordance, we removed the Quija 37 line from our mitochondrial and

565   *Wolbachia* data and computed Bayes Factors. The shared topology model was favored by a

566   Bayes factor of e^{55}, indicating strong support for concordance of the mitochondrial and

567   *Wolbachia* phylogenies.

568

569   **Variation in the ratio of mtDNA to *Wolbachia* substitution rates:** Following the methods of

570   Turelli *et al.* (2018) to look for different mitochondrial versus *Wolbachia* branch lengths, we also

571   found no evidence of variation in the ratio of mtDNA:*Wolbachia* substitution rates within the *D.*

572   *yakuba* clade (median 3rd position ratio $=$ 223, quartiles $=$ 186 and 268); the model with all

573   branches sharing the same ratio is favored over the model where each branch has its own ratio by

574   a Bayes factor of e^7 (Figure S4). For comparison, Turelli *et al.* (2018) found a median ratio of

575   566 within *D. suzukii* and 406 within *w*Ri-infected *D. ananassae*-subgroup species. These

576   differences in relative rates are broadly compatible with the tenfold variance in relative rates

577   noted above for nuclear versus mtDNA divergence across *Drosophila* species. Hence, our data

578   seem compatible with *Wolbachia* transmission via introgression within the *D. yakuba* clade.
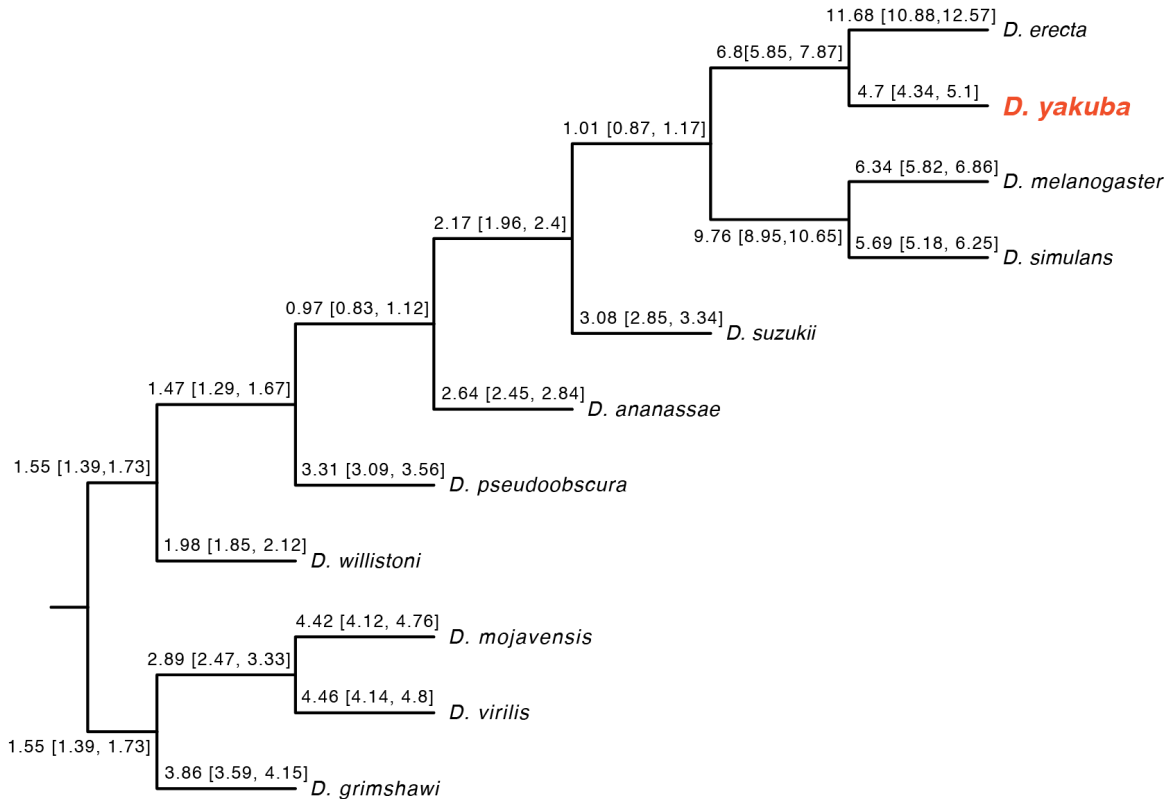
579

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 mitochondrial and nuclear phylograms are topologically identical when run separately.

580

581   Although this result seems to strongly support purely maternal transmission of *Wolbachia*

582   within species and introgressive transfer between species (with more recent introgression

583   between the sister species *D. yakuba* and *D. santomea*), this interpretation is severely weakened

584   by our "control" analysis that includes *w*Mel-infected *D. melanogaster* (for which introgression

585   is impossible). Including the *D. melanogaster* data, the constant-ratio model was still favored

586   over variable ratios by a Bayes factor of e^{54} (median 3rd position ratio = 297, quartiles = 276 and

587   323). We discuss the implications of this anomalous result below. The key observation is that

23

588 significantly declining rates of substitution for *Wolbachia* (and mtDNA, see Ho *et al*. 2005) over

589 time, together with heterogeneity of relative rates (as illustrated in Figure 4), limit the power of

590 relative substitution ratios to differentiate between introgression and horizontal transmission.

591

592 **Transposon-mediated transfer of CI factors independent of WO phage**

593 In contrast to *w*Mel, which contains only the *cidA-cidB* gene pair (LePage et al. 2017), the *D*.

594 *yakuba*-clade *Wolbachia*, which also belong to *Wolbachia* group A, have both *cidA-cidB* and a

595 *cinA-cinB* pair homologous to CI loci originally identified in group-B *w*Pip (Beckmann *et al*.

596 2017). The A and B *Wolbachia* groups diverged about 6-46 mya (Werren *et al*. 1995; Meany *et*

597 *al*. 2019). Among our *w*Yak-clade sequences, there are no single-nucleotide variants within any

598 of the *cif* loci. The *cidB*$^{w\text{Yak-clade}}$ locus has an inversion from amino acids 37–103 relative to the

599 same region in *cidB*$^{w\text{Mel}}$ in every *w*Yak-clade *Wolbachia* variant we analyzed. This introduces

600 several stop codons, which might render this gene nonfunctional. On the other hand, RNA

601 polymerase should still transcribe a complete polycistronic transcript. Therefore translation of an

602 N-terminally truncated CidB$^{w\text{Yak-clade}}$ protein cannot be ruled out. With the exception of a 236 bp

603 tandem duplication in *cinB*$^{w\text{Yak-clade}}$ (Figures 5 and 6), the sequence differences between *cinA*$^{w\text{Yak-}}$

604 $^{\text{clade}}$ and *cinB*$^{w\text{Yak-clade}}$ regions compared to *cinA*$^{w\text{Pip}}$ and *cinB*$^{w\text{Pip}}$ homologs are 1.35% and 0.59%,

605 respectively. In contrast, the average difference between *w*Yak and *w*Pip genomes across all 194

606 genes (161,655 bp) present in single copy and with identical lengths in both genomes is 11.60%.

607 Conversely, outside of the prophage regions, *w*Mel and *w*Yak differ by about 1%. This is

608 consistent with data that indicate WO phage regions have a different evolutionary history than

609 the bulk of the *Wolbachia* genome (LePage *et al*. 2017; Lindsey *et al*. 2018). The 236 bp tandem

610 duplication in the *cinB*$^{w\text{Yak-clade}}$ introduces a frame shift in the transcript at position 588. It is

611 unclear whether the *cinB*$^{w\text{Yak-clade}}$ protein retains functionality.

612 The absence of *cin* loci in *w*Mel, combined with the similarity of this region between *w*Yak

613 and distantly related *w*Pip, led us to assess how these *Wolbachia* acquired *cin* loci. Targeted

614 assembly extended the scaffold *w*Yak "702380" containing the *cin* loci (Figure 5) but did not

615 definitively place it relative to the *w*Mel genome. BLAST searches indicated that contig *w*Yak

616 "187383" was likely flanked by *w*Yak "702380". PCR primers designed in both contigs

617 amplified the intervening region (labeled "A" in Figure 5), confirming we have discovered the

618 first WO prophage with two sets of *cif* loci. Subsequent Sanger sequencing revealed that this

619 region contains an ISWpi1 element (Supplemental Information), found in many *Wolbachia*

620 genomes, but not present at this location in the *w*Mel reference sequence (Cordeaux *et al*. 2008).
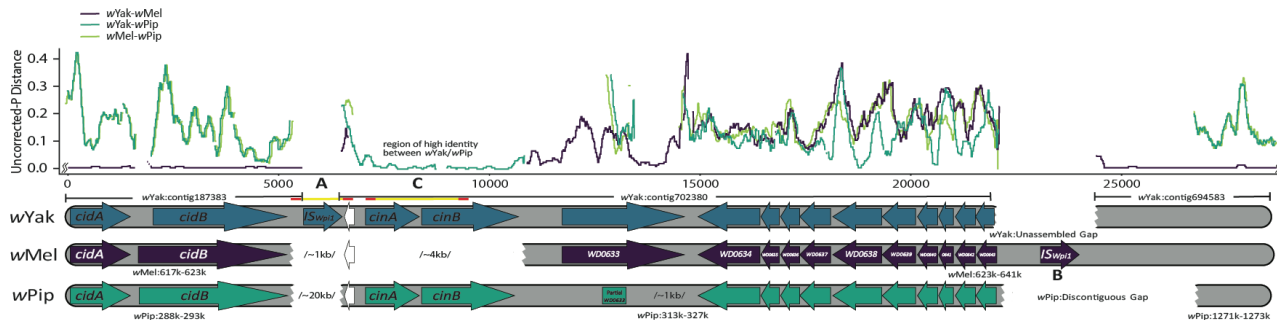
621

Figure 5 The gene structure and observed, uncorrected, pairwise distances among *w*Yak, *w*Mel, and *w*Pip *Wolbachia* spanning the region of the WO prophage island where the hypothesized transfer occurred between the ISWpi1 element "A" present in *w*Yak and the ISWpi1 sequence labeled "B" in *w*Mel. Over the region shown, *w*Yak, *w*San and *w*Tei are identical. The difference between *w*Yak and *w*Mel is less than 1% between the contigs (*w*Yak "187383" and *w*Yak "694583") flanking the region containing the *cinA-cinB* homologs. These homologs assembled on the first ~4,000 bp of the contig *w*Yak "702380", but they do not appear to exist in *w*Mel. A BLAST search against the NCBI nucleotide collection using contig *w*Yak "702380" as a query indicated the *cinA-cinB* homologs share high identity with group-B *w*Pip *Wolbachia* (~97%). The remainder of the *w*Yak contig is approximately equally divergent from both *w*Mel and *w*Pip, consistent with the entire region being transferred from an unknown *Wolbachia* variant that had acquired the *cinA-cinB* homologs. Under this scenario, ISWpi1 elements mediate the transfer to *w*Yak via excision of elements "A" and "B" along with the intervening DNA from the donor and subsequent homologous recombination between IS sequences "A" and "B" and WO phage island genes in *w*Yak. The ISWpi1 element "A" present in *w*Yak, but not in either *w*Mel or *w*Pip, was spanned using Sanger sequencing with primers anchored on unique sequence in the flanking contigs (Sanger sequences are denoted by yellow lines with primer locations denoted by red flanking regions, Table S1). Additionally, the presence of a 236 bp tandem duplication in *cinB* ("C") was confirmed using Sanger sequencing. The ISWpi1 sequence labeled "B" is assembled in *w*Mel and corresponds to *w*Mel#9 (Cordeaux 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).

622 IS elements encode a transposase gene that mediates their movement (Chandler and

623 Mahillon 2002). ISWpi1 elements are related to the IS5 family, which seems to be restricted to

624 *Wolbachia* (Duron *et al*. 2005; Cordeaux *et al*. 2008). ISWpi1 elements occur in more than half of

625 the *Wolbachia* strains that have been evaluated, including *w*Yak and *w*Mel (Cordeaux *et al*. 2008);

626 and these elements occur at variable copy number, potentially facilitating horizontal transmission

627 of ISWpi1 elements and their intervening sequence between *Wolbachia* variants/strains (Cordeaux

628 *et al*. 2008). Following placement of contig *w*Yak "702380" relative to *w*Mel, we aligned these

629 regions and calculated pairwise differences along the chromosome using a sliding window.

25

630   In addition to containing the *cinA-cinB* loci, absent in *w*Mel, contig *w*Yak "702380" is on

631   average about 10% different from *w*Mel. In contrast, most of *w*Yak is less than 1% different

632   from *w*Mel (for instance, across the 650,559 bp used in our phylogenetic analyses, *w*Mel and

633   *w*Yak differ by only 0.09%). Downstream of the *cinA-cinB*$^{w\text{Yak-clade}}$ region, targeted assembly

634   enabled us to join several more contigs. The junctions were corroborated by mapping paired-end

635   reads (Langmead and Salzberg 2012) and visually inspecting the resulting bam files around

636   joined contigs for reads spanning the new junctions and for concordant read-pair mappings.

637   However, our attempts to fully bridge the gap downstream of the *cinA-cinB*$^{w\text{Yak-clade}}$ genes via

638   targeted assembly, scaffolding, and PCR were unsuccessful (see "*w*Yak unassembled gap" in

639   Figure 5). The unassembled region in *w*Yak contains an ISWpi1 element in *w*Mel, Mel #9

640   (labeled "B" in Figure 5; Cordaux 2008). Although not part of our *w*Yak assembly, homologs of

641   this ISWpi1 element appear in assemblies of this region of the WO prophage in several other A-

642   group *Wolbachia* (Cordaux *et al.* 2008; Bordenstein and Bordenstein 2016), specifically *w*Inc

643   from *D. incompta* (Wallau *et al.* 2016) and *w*Ri from *D. simulans* (see Figure 2). In both, we find

644   orthologs to Mel#9 ISWpi1 with corresponding flanking sequence assembled, indicating that this

645   IS element probably occurs in this unassembled region of *w*Yak, *w*San, and *w*Tei. The *w*Yak

646   sequence between these two ISWpi1 elements is highly diverged from *w*Mel in comparison to

647   the rest of the *w*Yak genome (about 10% difference versus 1%). It therefore seems plausible that

648   this region experienced a horizontal transfer event in the ancestor of *w*Yak, *w*Tei, and *w*San,

649   mediated by the flanking ISWpi1 elements.

650   We conjectured that horizontal transfer occurred via the excision of the two ISWpi1

651   elements and the intervening DNA from the donor followed by homologous recombination

652   within the IS elements. To assess the plausibility of this scenario, we used BLAST to compare

653   *cinA-cinB*$^{w\text{Pip}}$ genes against all published *Wolbachia* genomes in the NCBI assembly database

654   (https://www.ncbi.nlm.nih.gov/assembly/?term=Wolbachia). We found homologs of the *cinA-*

655   *cinB*$^{w\text{Pip}}$ in the group-A *Wolbachia* associated with *Nomada* bees (Gerth and Bleidorn 2016). An

656   unrooted tree of the ORFs for *cinA* and *cinB* (Figure 6) indicates that these genes in *w*Yak, *w*San,

657   and *w*Tei are more similar to *cinA-cinB* from group-B *w*Pip and *w*AlbB than to *cinA-cinB* found

658   in the fellow group-A *Wolbachia* associated with *Nomada* bees (*w*NFe, *w*NPa, *w*NLeu, and

659   *w*NFa), which each harbor two different *cinA-cinB* copies. This indicates that *cinA-cinB*$^{w\text{Yak-clade}}$

660   were acquired via a horizontal transfer event across *Wolbachia* groups B and A that is

661   independent of the event(s) that placed *cinA-cinB* in the *Wolbachia* associated with *Nomada*

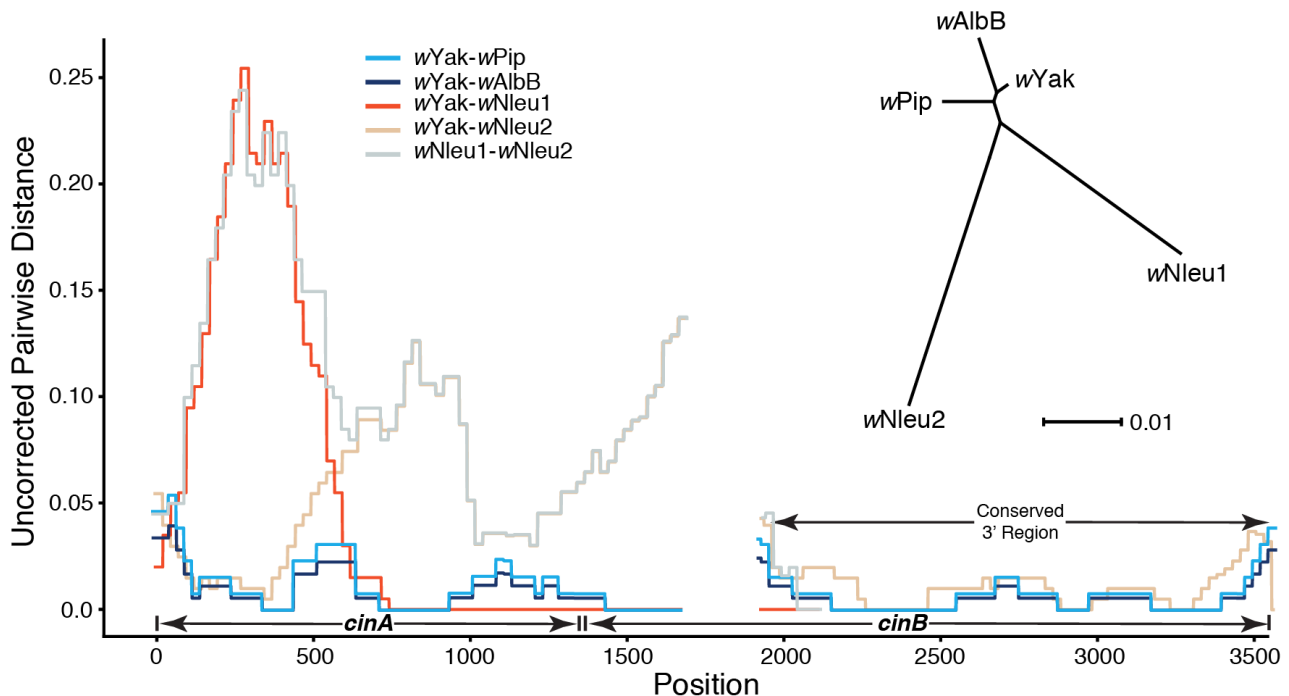662   bees, suggesting repeated transfers of *cin* loci.

663

664

Figure 6 The observed pairwise differences between *cinA* and *cinB* copies in *w*Yak-clade *Wolbachia* 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 this 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 gap at position 1705-1940 represents a tandem duplication present in *w*Yak. 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*).

665   The two *cinA-cinB* copies (denoted *w*NLeu1 and *w*NLeu2 in Figure 6) in *Nomada*

666   *Wolbachia* are nearly as distinct from each other as they are from the homologs in *w*Pip, *w*AlbB,

667   and *w*Yak (~7% diverged from these strains, Figure 6). However, among the four *Wolbachia*-

668   infected *Nomada* species, the orthologs are very similar, with $cinA^{wNLeu1}$ having only 0–0.15%

669   pairwise differences among the four strains and $cinA^{wNLeu2}$ having 0–0.56% pairwise differences.

670   (Reconstructions for *cinB* gene copies were more complicated as the $cinB^{wNLeu1}$ copy fails to

671   assemble into a single contig on the 3-prime end.) This pattern suggests that *w*NLeu1 and

672   *w*NLeu2 *cin* copies were acquired by the common ancestor of the four *Nomada Wolbachia*

673   strains analyzed, followed by cladogenic transfer across *Nomada* species (Gerth and Bleidorn

674   2016). The highly fragmented assemblies of the four *Nomada Wolbachia* strains, with duplicate

675   copies confounding assembly, make it difficult to determine the relative positions of the *cinA* and

676   *cinB* copies and if they are likewise flanked by ISWpi1 elements.

677   To determine whether these genes were potentially moved by ISWpi1 elements into the

678   *Wolbachia* of the *D. yakuba* clade and *Nomada*, we searched each genome using both the *cinA-*

679    $cinB^{wYak\text{-}clade}$ contig and the flanking ISWpi1 elements. Long repeated elements like ISWpi1

680   (916bp) break most short-read assemblies. Despite this, there is often a small fragment of the

681   element, the length of the short read, on either end of the broken contig, indicative of the repeat

682   element being responsible for terminating contig extension. We looked for the footprint of these

683   elements at the edges of the contigs on which the *cin* genes were found. We found ISWpi1

684   elements in the region flanking both $cinA^{wNLeu1}$ and $cinA^{wNLeu2}$ copies, consistent with our *w*Yak

685   assembly in which we verified the ISWpi1 element with Sanger sequencing. These data support

686   a role for ISWpi1 in the acquisition of the *cinA-cinB* genes by the *Wolbachia* in the *D. yakuba*

687   clade and the *Nomada* bees. We conjecture that future work will fully confirm ISWpi1 in the

688   horizontal movement of incompatibility loci between *Wolbachia*.

689

690   **DISCUSSION**

691   Our results indicate introgressive and horizontal *Wolbachia* acquisition in the *D. yakuba* clade.

692   Evidence for horizontal *Wolbachia* transfer here and elsewhere (Turelli *et al.* 2018) suggests that

693   double infections must be common, even if ephemeral. Such double infections provide an

694   opportunity for ISWpi1 transposable elements to mediate horizontal transfer of incompatibility

695   loci among divergent *Wolbachia*. Importantly, our results highlight that incompatibility factors

696   may move independently of prophage, as evidenced by our discovery of the first prophage

697   documented to have two sets of *cif* loci. We discuss these conclusions below.

698

699   **mtDNA Introgression**

700   Our relative mitochondrial chronogram provides strong support for three mitochondrial clades,

701   including a monophyletic *D. teissieri* clade that is outgroup to two sister clades: one clade

702   consisting of mitochondria from 14 *D. santomea* individuals, and the other contains all *D.*

703   *yakuba* mitochondria plus mitochondria from 12 *D. santomea* individuals. Our results suggest

704   less mitochondrial introgression in the *D. yakuba* clade than a past report that used sequence data

705   from only two mitochondrial loci (*COII* and *ND5*, 1777 bp) and reported a clade with

706   mitochondria sampled from each species represented (Figure 3 in Bachtrog *et al*. 2006). (Using

707   sequence data from only *COII* and *ND5*, we can replicate this result, indicating data from

708   additional loci are needed to add resolution.) Our results also agree with recent work that

709   demonstrated little nuclear introgression in the *D. yakuba* clade (Turissini and Matute 2017).

710   Our results broadly agree with the results produced by Llopart *et al.* (2014) who assessed

711   mitochondrial introgression between *D. santomea* and *D. yakuba* using whole mitochondrial

712   genomes. They generated a neighbor-joining tree that produced a clade consisting of all *D.*

713   *yakuba* individuals and 10 *D. santomea* individuals, and this clade is sister to a clade with 6 *D.*

714   *santomea* individuals; one *D. santomea* haplotype is outgroup to all other haplotypes included in

715   their analysis (Figure 3 in Llopart *et al.* 2014). Nested within the mixed *D. yakuba* clade, Llopart

716   *et al.* (2014) identified a "hybrid zone clade" that includes *D. yakuba* individuals sampled from

717   São Tomé and from continental Africa and *D. santomea* individuals sampled from both within

718   and outside the Pico de São Tomé hybrid zone (HZ). The sister *D. santomea* clade also contains

719   both HZ and non-HZ individuals. Thus, their analysis and ours provide support for hybridization

720   within and outside of the HZ, leading us to question both the existence of a HZ clade and the

721   claim that these species "share the same mitochondrial genome" (Llopart *et al.* 2014). Instead,

722   both our results and theirs suggest unidirectional introgression of *D. yakuba* mitochondria into *D.*

723   *santomea*; we find 59% of *D. santomea* individuals having *D. yakuba*-like mitochondria and they

724   found 46%.

725   Llopart *et al.* (2014) used a strict molecular clock to estimate the mitochondrial MRCA of *D.*

726   *santomea* and *D. yakuba* at 10,792–17,888 years by calibrating their tree with the *D. yakuba-D.*

29

727   *erecta* split, estimated at 10.4 mya (Tamura *et al.* 2004). A high level of mitochondrial saturation

728   over time, with an expected value of 1.44 substitutions per synonymous site for the *D. yakuba-D.*

729   *erecta* split, could influence this estimate (Llopart *et al.* 2014). Moreover, Ho *et al.* (2005)

730   demonstrated that the mtDNA substitution rate resembles an exponential curve, with high short-

731   term substitution rates that approach the mutation rate, then slowing to a long-term rate after

732   about 1–2 million years of divergence, far younger than the inferred *D. yakuba-D. erecta* nuclear

733   and mtDNA divergence. Hence, using the slow long-term *D. yakuba-D. erecta* calibration is

734   likely to underestimate the more rapid rates of divergence experienced by *D. yakuba*-clade

735   mtDNA, inflating divergence-time estimates (Ho *et al.* 2005). If we assume that *Wolbachia* and

736   mitochondria were transferred by introgression, which our analyses support, our estimates in

737   Figure 3 suggest that *D. santomea* and *D. yakuba* mtDNA diverged more recently, with our point

738   estimates ranging from about 1,500 to 2,800 years.

739

740   **_Wolbachia_ placement, divergence, and acquisition**

741   **_Wolbachia_ placement:** Despite their similarity, *w*San, *w*Tei, and *w*Yak form monophyletic

742   clades with *w*Tei outgroup to sisters *w*San and *w*Yak, recapitulating host relationships (Figure 1;

743   Turissini and Matute 2017). *w*Mel is sister to the *D. yakuba*-clade strains in the A group (Figure

744   2), which also includes *D. simulans* strains (*w*Ri, *w*Au, and *w*Ha), *w*Ana, *w*Suz, and the *Nomada*

745   *Wolbachia* (*w*NFa, *w*NLeu, *w*NPa, and *w*NFe). Our phylogram (Figure 2A) places *w*AlbB

746   outgroup to *w*No and *w*Pip strains that diverged from A-group *Wolbachia* about 6–46 mya

747   (Meany *et al.* 2019).

748

749   **_Wolbachia_ divergence:** Our chronogram analyses (Figure 3) estimate that *D. yakuba*-clade

750   *Wolbachia* and three *w*Mel variants diverged about 29,000 years ago, and that *w*Tei split from

751   sisters *w*San and *w*Yak about 2,500–4,500 years ago, with *w*San and *w*Yak diverging about

752   1,600-2,800 years ago. We estimate that the two most divergent *w*Mel variants from Richardson

753   *et al.* (2012) and the reference *w*Mel genome split about 4,900–7,200 years ago, indicating more

754   divergence among *w*Mel variants than among *D. yakuba*-clade *Wolbachia* strains. All of these

755   results depend on the calibration provided by Richardson *et al.* (2012) and the relative accuracy

756   of the underlying models of molecular evolution, which assume constant relative rates of change

757   across data partitions. For the deepest divergence in Figure 3, between *D. yakuba*-clade

758   *Wolbachia* and *w*Mel, we find that the estimated time depends on the variance in our prior

759   distribution for substitution rates across branches, with a strict clock putting the divergence at

760   about 73,000 years rather than 29,000 years obtained with the most variable prior. Despite this

761   uncertainty, the quantitative differences of our *Wolbachia* divergence-time estimates do not alter

762   the qualitative conclusion that these *Wolbachia* did not co-diverge with these hosts, which split

763   several million years ago.

764   Our findings here and in Turelli *et al*. (2018) suggest that for several *Drosophila* species,

765   their current *Wolbachia* infections have been in residence for only hundreds to tens of thousands

766   of years. Bailly-Bechet *et al*. (2017) estimated *Wolbachia* residence times using data from more

767   than 10,000 arthropod specimens spanning over 1,000 species. However, they analyzed DNA

768   sequences from only part of the fast-evolving *fbpA Wolbachia* locus and the host *CO1* mtDNA

769   locus. From an initial model-based meta-analysis, they concluded that "... most infections are

770   very recent ..." – consistent with our results. However, they also fit a more complex model with

771   "short" and "long" time scales for acquisition and loss, conjecturing that short-term rates were

772   associated with imperfect maternal transmission. Focusing on long-time rates, they concluded

773   that *Wolbachia* infections persisted in lineages for approximately 7 million years on average,

774   whereas lineages remained uninfected for about 9 million years. For *Drosophila*, such long

775   infection durations would imply that *Wolbachia*-host associations often persist through

776   speciation (see Coyne and Orr 1997 and Turelli *et al*. 2014 for estimates of speciation times in

777   *Drosophila*, generally 10^5–10^6 years). Extrapolating from limited *Wolbachia* sequence data,

778   Hamm *et al*. (2014) conjectured that cladogenic *Wolbachia* transmission might be common

779   among *Drosophila*; but this extrapolation is refuted by our genomic analyses. The long

780   *Wolbachia* durations proposed Bailly-Bechet *et al*. (2017) depend on their conjecture that their

781   long-term rate estimates accurately reflect acquisition and loss of *Wolbachia* infections across

782   species. This is worth testing with additional analyses of *Wolbachia*, mitochondrial and nuclear

783   genomes from a broad range of arthropods.

784

785   ***Wolbachia* acquisition—introgression versus horizontal:** Our divergence-time estimates for

786   the *D. yakuba*-clade *Wolbachia* versus their hosts preclude cladogenic acquisition. Unlike

787   introgression, horizontal (or paternal) *Wolbachia* acquisition should produce discordance

788   between phylogenies inferred for *Wolbachia* and the associated mitochondria. With the notable

789   exception of *D. santomea* line Quija 37, which has mitochondria belonging to the clade

790   associated with *D. yakuba*, but has *w*San *Wolbachia*, we find no evidence of discordance

791   between the estimated mitochondrial and *Wolbachia* phylogenies. Hence our data indicate that

792   acquisition by introgression is far more common than horizontal transmission between closely

793   related species, consistent with data on acquisition of *w*Ri-like *Wolbachia* in the *D. melanogaster*

794   species group (Turelli *et al*. 2018). Similarly, consistent with extensive data from *D.*

795   *melanogaster* (Richardson *et al*. 2012) and *D. simulans* (Turelli *et al*. 2018) and smaller samples

796   from *D. suzukii* and *D. ananassae* (Turelli *et al.* 2018), we find only one possible example of

797   paternal transmission or horizontal transmission within *D. yakuba*-clade species.

798   We also investigated an alternative approach to distinguish between introgressive and

799   horizontal *Wolbachia* acquisition by estimating substitution ratios for mtDNA versus *Wolbachia*

800   (Turelli *et al.* 2018). Because we could estimate this ratio on each branch, we conjectured that

801   this approach might have greater resolving power than our incompletely resolved mitochondrial

802   and *Wolbachia* phylogenies. We expect higher ratios with horizontal transmission because

803   mtDNA would have been diverging longer than recently transferred *Wolbachia*. However, this

804   approach assumes that mtDNA and *Wolbachia* substitution rates remain relatively constant. This

805   is contradicted by the finding of Ho *et al.* (2005) that mtDNA substitution rates decline

806   substantially with increasing divergence time, reaching an asymptote after around 1–2 million

807   years. To calibrate their *Wolbachia* substitution rate estimates, Richardson *et al.* (2012) used an

808   experimentally observed mitochondrial mutation rate in *D. melanogaster* (6.2×10^{-8} mutations

809   per third-position site per generation) that extrapolates to 62% third-position divergence per

810   million years. This extrapolation is nonsensical as a long-term substitution rate. As summarized

811   by Ho *et al*. (2005), typical mtDNA substitution rates are 0.5–1.5% per coding site per million

812   years. Nevertheless, using the ratio of short-term rates for mtDNA and *Wolbachia*, Richardson *et*

813   *al.* (2012) produced an estimate of the long-term *Wolbachia* substitution rate that agrees with

814   independent estimates from *Nasonia* wasps (Raychoudhury *et al.* 2009) and *Nomada* bees (Gerth

815   and Bleidorn 2016) derived from much longer divergence times (assuming cladogenic

816   *Wolbachia* acquisition) (Conner *et al*. 2017). This paradox is resolved if *Wolbachia* molecular

817   evolution is not subject to the dramatic slowdown in rates seen for mtDNA substitution rates.

818   The apparent difference between mtDNA molecular evolution (dramatic slowdown over

819   longer time scales, Ho *et al*. 2005) and *Wolbachia* molecular evolution (relative constancy, as

820 inferred from similar rates of differentiation over very different time scales) suggests why our

821 relative-rate test does not reject introgressive transmission of *Wolbachia* between *D.*

822 *melanogaster* and the *D. yakuba* clade, even though it is clearly impossible. A roughly 50-fold

823 slowdown in mtDNA substitution rates over the time scale of the divergence of *D. melanogaster*

824 from the *D. yakuba* clade, relative to the rate of differentiation within the *D. yakuba* clade,

825 produces comparable mtDNA-*Wolbachia* substitution ratios for comparisons within the *D.*

826 *yakuba* clade and between *D. melanogaster* and the *D. yakuba* clade. Because of this

827 complication and our conjecture that relative rates of mtDNA versus *Wolbachia* substitutions

828 over longer periods are likely to mirror the ten-fold differences we see for mtDNA versus

829 nuclear genes, phylogenetic discordance between mitochondria and *Wolbachia* is clearly a much

830 more robust indicator of horizontal (or paternal) *Wolbachia* transmission. Nevertheless,

831 additional examples of cladogenic *Wolbachia* acquisition are needed to better understand relative

832 rates and patterns of *Wolbachia*, mtDNA and nuclear differentiation over different time scales.

833 Our divergence-times estimate of the *D. yakuba*-clade *Wolbachia* versus their hosts precludes

834 cladogenic transmission; and our phylogenetic analyses suggest that these species share very

835 similar *Wolbachia* because of introgression, as originally argued by Lachaise *et al.* (2000).

836 However, under either introgressive or horizontal transfer of *Wolbachia*, we expect the donor

837 species *Wolbachia* sequences would appear paraphyletic when analyzed jointly with the

838 *Wolbachia* from the recipient. Paraphyly allowed Turelli *et al.* (2018) to infer that *D. simulans*

839 likely obtained its *Wolbachia* from *D. ananassae* and *D. subpulchrella* likely obtained its

840 *Wolbachia* from *D. suzukii*. Paraphyly is generally expected soon after gene flow stops between

841 populations. As noted by Hudson and Coyne (2002, Fig. 1), the time scale expected to produce

842 reciprocal monophyly for mitochondria (and *Wolbachia*) under a neutral model of molecular

843 evolution is on the order of the effective size of the species. Our results in Figure 3 indicate that,

844 at least for our small samples, reciprocal monophyly for the *Wolbachia* in these three species has

845 been achieved within a few thousand years. This suggests that reciprocal monophyly has been

846 accelerated by species-specific selective sweeps within the *Wolbachia* or mitochondria of these

847 species. This conjecture may be testable from estimates of host fitness using transinfected versus

848 native *Wolbachia*.

849

IS transposable elements mediate horizontal transfer of incompatibility loci between divergent *Wolbachia*

Wolbachia in all three *D. yakuba*-clade hosts cause both intra-and interspecific CI (Cooper *et al.* 2017), despite originally being characterized as non-CI causing (Zabalou *et al.* 2004; Charlat *et al.* 2004). CI is relatively weak, and its strength can vary among *w*Tei variants and *D. teissieri* backgrounds (Table 3 in Cooper *et al.* 2017). Differences in CI among *Wolbachia* variants has also been demonstrated in interspecific backgrounds where *w*Tei caused stronger CI in a *D. simulans* background (97.2 ± 1.3 SE percent embryo mortality) than either *w*Yak (26.5 ± 4.2 SE, percent embryo mortality) or *w*San (24.0 ± 4.1 SE, percent embryo mortality) (Zabalou et al. 2008). Both *w*Yak and *w*San induced CI in *D. simulans* comparable in intensity to that found by Cooper *et al.* (2017, Figure 3) in their original hosts. Surprisingly, Zabalou *et al*. (2008) found that the strength of CI induced by *w*Tei in *D. simulans* even eclipsed that of *w*Ri (89.8 ± 4.5 SE percent embryo mortality). These results must be reconciled with the fact that loci known to underlie CI do not vary within or among *D. yakuba*-clade *Wolbachia* variants we examined. The nearly complete CI induced by *w*Tei in *D. simulans* may depend on CI-causing factors yet to be identified or differences in gene expression.

In each *D*. *yakuba*-clade *Wolbachia* variant included in our analyses, we find a disruption of *cidB*$^{w\text{Yak-clade}}$ with an inversion from amino acids 37–103 relative to the same region in sister *w*Mel. The inversion introduces multiple stop codons that could render this gene nonfunctional. Fixation of loss-of-function mutations in CI-causing loci is consistent with theoretical analyses showing that selection on *Wolbachia* within host lineages does not act to increase or maintain CI (Prout 1994; Turelli 1994; Haygood and Turelli 2009); indeed, we have also recently observed a single mutation that disrupts *cidB* in non-CI causing *w*Mau *Wolbachia* that infect *D*. *mauritiana* on the island of Mauritius (Meany et al. 2019). In both *w*Mau and the *D*. *yakuba*-clade *Wolbachia*, we find fixation of defects in the putative toxin gene. We expect that future genomic analyses will produce additional examples.

All *D*. *yakuba*-clade *Wolbachia* genomes included in our analysis harbor *cinA-cinB* loci originally discovered in the *w*Pip strain that diverged from A-group *Wolbachia*, including the *D*. *yakuba*-clade variants, about 6–46 mya (Meany *et al*. 2019). *cin* loci are also present in B-group *w*AlbB that infects *Ae*. *albopictus* and in A-group *w*NFe, *w*NPa, *w*NLeu, and *w*NFa *Wolbachia* that infect *Nomada* bees. *cin* loci are absent from *w*Mel, but the *w*Yak contig containing these

881   loci is about 10% diverged from *w*Mel, while observed divergence between *w*Yak and *w*Mel

882   across the rest of the genome is less than 1%. The *w*Yak-clade *cin* loci share about 97%

883   similarity with the divergent B-group *w*Pip strain. *w*Yak-clade *cin* loci are more similar to *cinA-*

884   *cinB* from the B-group *Wolbachia* *w*Pip and *w*AlbB than to those in A-group *Nomada Wolbachia*

885   strains, which have two sets of *cin* loci that are as diverged from each other as they are from

886   these regions in *w*Yak and in B-group *Wolbachia*. These observations suggest independent

887   horizontal transfer of *cin* loci into *w*Yak and *Nomada Wolbachia*.

888   Our results indicate that independent of prophage movement, ISWpi1-element paralogs can

889   move incompatibility loci via the excision of flanking ISWpi1 elements, followed by

890   homologous recombination within the elements. Horizontal *Wolbachia* acquisition is common in

891   *Drosophila* (Turelli *et al.* 2018) and other species (O'Neill *et al.* 1992), suggesting that double

892   infections, which could provide the opportunity for ISWpi1-mediated transfer of incompatibility

893   loci, may be common, even if transient. (A second infection need not become stably transmitted

894   for horizontal gene transfer via ISWpi1 elements to occur.) In contrast, phage particles or virions

895   could be introduced by a vector and provide the opportunity for ISWpi1 mediated transfer

896   (Ahmed *et al.* 2015; Brown and Lloyd 2015), without the presence of a double *Wolbachia*

897   infection. Determining whether the insertion of these loci was derived from a prophage region of

898   the *Wolbachia* genome, or from a phage genome encapsulated in a phage particle, remains an

899   open question. While the ISWpi1 element in *w*Mel (Mel #9 labeled "B" in Figure 5; Cordaux

900   2008) is not part of our *w*Yak assembly, homologs of this element are present in assemblies of

901   several other A-group *Wolbachia* including *w*Inc and *w*Ri (Cordaux *et al.* 2008; Bordenstein and

902   Bordenstein 2016). We predict this element occurs in the unassembled region of our *w*Yak

903   assembly. Footprints of ISWpi1 elements in the region flanking the *cinA* genes for both copies of

904   the gene in the *Nomada Wolbachia* provide further support for our hypothesis. Long-read-based

905   *Wolbachia* assemblies from many infected host systems will elucidate the role of ISWpi1

906   elements in horizontal transfer of CI loci. Overall, the ecology of horizontal *Wolbachia*

907   transmission is crucial to understanding *Wolbachia* acquisition; and the transfer and dynamics of

908   CI loci are crucial to understanding *Wolbachia* evolution.

909

910

911

LITERATURE CITED

Ahmed, M. Z., S.-J. Li, X. Xue, X.-J. Yin, S.-X. Ren *et al.*, 2015 The intracellular bacterium *Wolbachia* uses parasitoid wasps as phoretic vectors for efficient horizontal transmission. PLoS Pathogens 11**:** e1004672.

Bachtrog, D., K. Thornton, A. Clark and P. Andolfatto, 2006 Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. Evolution 60**:** 292-302.

Bailly-Bechet, M., P. Martins-Simoes, G. J. Szollosi, G. Mialdea, M.-F. Sagot *et al.*, 2017 How long does *Wolbachia* remain on board. Mol. Biol. Evol. 34: 1183–1193.

Bakovic V., M. Schebeck, A. Telschow, C. Stauffer, H. Schuler, 2018 Spatial spread of *Wolbachia* in *Rhagoletis cerasi* populations. Biol. Lett. 14: 20180161

Baldo, L., J. C. D. Hotopp, K. A. Jolley, S. R. Bordenstein, S. A. Biber *et al.*, 2006 Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. Appl. Environ. Microbiol. 72**:** 7098-7110.

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

Bandi, C., T. J. Anderson, C. Genchi and M. L. Blaxter, 1998 Phylogeny of *Wolbachia* in filarial nematodes. Proc. Biol. Sci. 265**:** 2407–2413.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin *et al.*, 2012 SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19**:** 455–477.

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

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

Beckmann, J. F., M. Bonneau, H. Chen, M. Hochstrasser, D. Poinsot *et al.*, 2019 The toxin-antidote model of cytoplasmic incompatibility: genetics and evolutionary implications. Trends Genet. 35: 175–185.

Bordenstein, S. R., and S. R. Bordenstein, 2016 Eukaryotic association module in phage WO genomes from *Wolbachia*. Nat. Commun. 7**:** 13155.

Bouchon, D., T. Rigaud and P. Juchault, 1998 Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. Proc. Biol. Sci. 265: 1081–1090.

Braig, H. R., W. Zhou, S. L. Dobson and S. L. O'Neill, 1998 Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. J. Bacteriol. 180: 2373–2378.

Brand, C. L., S. B. Kingan, L. Wu and D. Garrigan, 2013 A selective sweep across species 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.

Cattel, J., K. Nikolouli, T. Andrieux, J. Martinez, F. Jiggins, S. Charlat, F. Vavre, D. Lejon, P. Gibert, and L. Mouton. 2018. Back and forth *Wolbachia* transfers reveal efficient strains to control spotted wing *Drosophila* populations. Journal of Applied Ecology 55:2408–2418.

Chandler, M., and J. Mahillon, 2002Insertion Sequences Revisited, p 305-366. *In* Craig N, Craigie R, Gellert M, Lambowitz A (ed), *Mobile DNA II*. ASM Press, Washington, DC. doi: 10.1128/9781555817954.ch15.

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.

Choi, J. Y., J. E. Bubnell, C. F. Aquadro, 2015 Population genomics of infectious and integrated *Wolbachia pipientis* genomes in *Drosophila ananassae*. Genome Biology and Evolution 7**:** 2362–2382.

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. 7**:** 9391–9404.

Cooper, B. S., P. S. Ginsberg, M. Turelli and D. R. Matute, 2017 *Wolbachia* in the *Drosophila yakuba* complex: pervasive frequency variation and weak cytoplasmic incompatibility, but no apparent effect on reproductive isolation. Genetics 205**:** 333–351.

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.

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., 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.

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.

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

Ellegaard, K. M., L. Klasson, K. Näslund, K. Bourtzis and S. G. E. Andersson, 2013 Comparative genomics of *Wolbachia* and the bacterial species concept. PLoS Genetics 9**:** e1003381.

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

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

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

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

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

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

Havird, J. C., and D. B. Sloan, 2016 The roles of mutation, selection, and expression in determining relative rates of evolution in mitochondrial versus nuclear genomes. Mol. Biol. Evol. 33**:** 3042-3053.

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

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

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

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

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

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

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

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

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., 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., 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.

Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot *et al.*, 2016 RevBayes: bayesian phylogenetic inference using graphical models and an interactive model-specification language. Syst. Biol. 65: 726-736.

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

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

Huigens, M. E., R. P. de Almeida, P. A. Boons, R. F. Luck and R. Stouthamer, 2004 Natural interspecific and intraspecific horizontal transfer of parthenogenesis-inducing *Wolbachia* in Trichogramma wasps. Proc. Biol. Sci. 271**:** 509-515.

Huigens, M. E., R. F. Luck, R. H. Klaassen, M. F. Maas, M. J. Timmermans *et al.*, 2000 Infectious parthenogenesis. Nature 405**:** 178-179.

Hunter, S. S., R. T. Lyon, B. A. J. Sarver, K. Hardwick, L. J. Forney *et al.*, 2015 Assembly by Reduced Complexity (ARC): a hybrid approach for targeted assembly of homologous sequences. bioRxiv**:** 014662.

Hurst, G. D. D., and F. M. Jiggins, 2000 Male-killing bacteria in insects: mechanisms, incidence, and implications. Emerg. Infect. Dis. 6**:** 329-336.

Jackman, S. D., B. P. Vandervalk, H. Mohamadi, J. Chu, S. Yeo *et al.*, 2017 ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome Res. 27**:** 768-777.

Jaenike, J., K. A. Dyer, C. Cornish and M. S. Minhas, 2006 Asymmetrical reinforcement and *Wolbachia* infection in *Drosophila*. PLoS Biol. 4**:** 1852-1862.

Joshi, N. A., and J. N. Fass, 2011 Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files, pp.

Kamimura, Y., and H. Mitsumoto, 2012 Lock-and-key structural isolation between sibling *Drosophila* species. Entomol. Sci. 15**:** 197-201.

Katoh, K., and D. M. Standley, 2013 MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30**:** 772-780.

Kern, A. D., D. A. Barbash, J. Chang Mell, A. Jensen and D. Hupalo, 2015 Highly constrained intergenic *Drosophila* ultraconserved elements are candidate ncRNAs. Genome Biol. Evol. 7**:** 689-698.

Klasson, L., T. Walker, M. Sebaihia, M. J. Sanders, M. A. Quail *et al.*, 2008 Genome evolution of *Wolbachia* strain *w*Pip from the *Culex pipiens* group. Mol. Biol. Evol. 25**:** 1877-1887.

Klasson, L., J. Westberg, P. Sapountzis, K. Näslund, Y. Lutnaes *et al.*, 2009 The mosaic genome structure of the *Wolbachia w*Ri strain infecting *Drosophila simulans*. PNAS 106**:** 5725-5730.

Kondo, R., Y. Satta, E. T. Matsuura, H. Ishiwa, N. Takahata *et al.*, 1990 Incomplete maternal transmission of mitochondrial DNA in *Drosophila*. Genetics 126**:** 657-663.

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

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

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

Kulathinal, R. J., L. S. Stevison and M. A. F. Noor, 2009 The genomics of speciation in *Drosophila*: diversity, divergence, and introgression estimated using low-coverage genome sequencing. PLoS Genet. 5**:** e1000550.

Lachaise, D., M. Harry, M. Solignac, F. Lemeunier, V. Benassi *et al.*, 2000 Evolutionary novelties in islands: *Drosophila santomea*, a new *melanogaster* sister species from São Tomé. Proc. Sci. Biol. 267**:** 1487-1495.

Langley, C. H., and W. M. Fitch, 1974 An examination of the constancy of the rate of molecular evolution. J. Mol. Evol. 3**:** 161-177.

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nature Methods 9**:** 357-359.

Laven, H., 1951 Crossing experiments with Culex strains. Evolution 5**:** 370-375.

Laven, H., 1967 Eradication of *Culex pipiens fatigans* through cytoplasmic incompatibility. Nature 216**:** 383-384.

LePage, D. P., J. A. Metcalf, S. R. Bordenstein, J. On, J. I. Perlmutter *et al.*, 2017 Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. Nature 543**:** 243-247.

Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27**:** 2987-2993.

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

Li, S.-J., M. Z. Ahmed, N. Lv, P.-Q. Shi, X.-M. Wang *et al.*, 2017 Plantmediated horizontal transmission of *Wolbachia* between whiteflies. ISME Journal 11**:** 1019-1028.

Lindsey, A. R. I., D. W. Rice, S. R. Bordenstein, A. W. Brooks, S. R. Bordenstein *et al.*, 2018 Evolutionary Genetics of Cytoplasmic Incompatibility Genes *cifA* and *cifB* in Prophage WO of *Wolbachia*. Genome Biol. Evol. 10**:** 434-451.

Llopart, A., D. Herrig, E. Brud and Z. Stecklein, 2014 Sequential adaptive introgression of the mitochondrial genome in *Drosophila yakuba* and *Drosophila santomea*. Mol. Ecol. 23**:** 1124-1136.

Lohse, K., M. Clarke, M. G. Ritchie and W. J. Etges, 2015 Genome-wide tests for introgression between cactophilic *Drosophila* implicate a role of inversions during speciation. Evolution 69**:** 1178-1190.

Martinez, J., B. Longdon, S. Bauer, Y.-S. Chan, W. J. Miller *et al.*, 2014 Symbionts commonly provide broad spectrum resistance to viruses in insects: a comparative analysis of *Wolbachia* strains. PLoS Path. 10**:** e1004369.

Matute, D. R., 2010 Reinforcement of gametic isolation in *Drosophila*. PLoS Biol. 8**:** e1000341.

Matute, D. R., and J. F. Ayroles, 2014 Hybridization occurs between *Drosophila simulans* and *D. sechellia* in the Seychelles archipelago. J. Evol. Biol. 27**:** 1057-1068.

Matute, D. R., and J. A. Coyne, 2010 Intrinsic reproductive isolation between two sister species of *Drosophila*. Evolution 64**:** 903-920.

Mavingui, P., C. Valiente Moro, V. Tran-Van, F. Wisniewski-Dye, V. Raquin *et al.*, 2012 Whole-genome sequence of *Wolbachia* strain *w*AlbB, an endosymbiont of tiger mosquito vector *Aedes albopictus*. J. Bacteriol. 194**:** 1840.

McMeniman, C. J., R. V. Lane, B. N. Cass, A. W. C. Fong, M. Sidhu *et al.*, 2009 Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti*. Science 323**:** 141-144.

Meany, M. K., W. R. Conner, S. V. Richter, J. A. Bailey, M. Turelli *et al.*, in press Loss of cytoplasmic incompatibility and minimal fecundity effects explain relatively low *Wolbachia* frequencies in *Drosophila mauritiana*. Evolution. doi: 10.1111/evo.13745.

Moriyama, M., N. Nikoh, T. Hosokawa and T. Fukatsu, 2015 Riboflavin provisioning underlies *Wolbachia*'s fitness contribution to its insect host. mBio 6. e01732-15.

Nosil, P., 2013 Degree of sympatry affects reinforcement in *Drosophila*. Evolution 67**:** 868-872.

Nurk, S., A. Bankevich, D. Antipov, A. Gurevich, A. Korobeynikov *et al.*, 2013 Assembling genomes and mini-metagenomes from highly chimeric reads, pp. 158-170 in *Research in Computational Molecular Biology*, edited by M. Deng, R. Jiang, F. Sun and X. Zhang. Springer Berlin Heidelberg, Berlin, Heidelberg.

O'Connor, L., C. Plichart, A. C. Sang, C. L. Brelsfoard, H. C. Bossin *et al.*, 2012 Open release of male mosquitoes infected with a *Wolbachia* biopesticide: field performance and infection containment. PLoS Negl. Trop. Dis. 6**:** e1797.

O'Neill, S. L., R. Giordano, A. M. Colbert, T. L. Karr and H. M. Robertson, 1992 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. PNAS 89**:** 2699-2702.

O'Neill, S. L., A. A. Hoffman and J. H. Werren, 1998 *Influential Passengers: Inherited Microorganisms and Arthropod Reproduction*. Oxford University Press.

Prout, T., 1994 Some evolutionary possibilities for a microbe that causes incompatibility in its host. Evolution 48**:** 909-911.

Raychoudhury, R., L. Baldo, D. C. S. G. Oliveira and J. H. Werren, 2009 Modes of acquisition of *Wolbachia*: horizontal transfer, hybrid introgression, and codivergence in the *Nasonia* species complex. Evolution 63**:** 165-183.

Richardson, M. F., L. A. Weinert, J. J. Welch, R. S. Linheiro, M. M. Magwire *et al.*, 2012 Population genomics of the *Wolbachia* endosymbiont in *Drosophila melanogaster*. PLoS Genet. 8**:** e1003129.

Ritchie S. A., 2018 *Wolbachia* and the near cessation of dengue outbreaks in Northern Australia despite continued dengue importations via travellers. J. Travel Med. 25: tay084.

Rousset, F., and M. Solignac, 1995 Evolution of single and double *Wolbachia* symbioses during speciation in the *Drosophila simulans* complex. PNAS 92**:** 6389-6393.

Rousset, F., D. Bouchon, B. Pintureau, P. Juchault and M. Solignac, 1992 *Wolbachia* endosymbionts responsible for various alterations of sexuality in arthropods. Proc. Biol. Sci. 250: 91-98.

Sánchez, L., and P. Santamaria, 1997 Reproductive isolation and morphogenetic evolution in *Drosophila* analyzed by breakage of ethological barriers. Genetics 147**:** 231-242.

Schmidt, T. L., N. H. Barton, G. Rašić, A. P. Turley, B. L. Montgomery *et al.*, 2017 Local introduction and heterogeneous spatial spread of dengue-suppressing *Wolbachia* through an urban population of *Aedes aegypti*. PLoS Biol. 15**:** e2001894.

Schrider, D. R., J. Ayroles, D. R. Matute and A. D. Kern, 2018 Supervised machine learning reveals introgressed loci in the genomes of *Drosophila simulans* and *D. sechellia*. PLoS Genet. 14**:** e1007341.

Schuler, H., K. Köppler, S. Daxböck-Horvath, B. Rasool, S. Krumböck *et al.*, 2016 The hitchhiker's guide to Europe: the infection dynamics of an ongoing *Wolbachia* invasion and mitochondrial selective sweep in *Rhagoletis cerasi*. Mol. Ecol. 25**:** 1595-1609.

Seemann, T., 2014 Prokka: rapid prokaryotic genome annotation. Bioinformatics 30**:** 2068-2069.

Shoemaker, D. D., V. Katju and J. Jaenike, 1999 *Wolbachia* and the evolution of reproductive isolation between *Drosophilla recens* and *Drosophila subquinaria*. Evolution 53**:** 1157-1164.

Shropshire, J. D., J. On, E. M. Layton, H. Zhou and S. R. Bordenstein, 2018 One prophage WO gene rescues cytoplasmic incompatibility in *Drosophila melanogaster*. PNAS 115: 4987-4991.

Shropshire, J. D., B. Leigh, S. R. Bordenstein, A. Duplouy, M. Riegler et al., 2019 Models and nomenclature for cytoplasmic incompatibility: Caution over premature conclusions – a response to Beckmann *et al.* Trends in Genetics 35: 397-399.

Simao, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva and E. M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31**:** 3210-3212.

Siozios, S., A. Cestaro, R. Kaur, I. Pertot, O. Rota-Stabelli, and G. Anfora, 2013 Draft genome sequence of the *Wolbachia* endosymbiont of *Drosophila suzukii*. Genome Announcements. 1, e00032-13.

Stamatakis, A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics (Oxford, England) 30**:** 1312-1313.

Sutton, E. R., S. R. Harris, J. Parkhill and S. P. Sinkins, 2014 Comparative genome analysis of *Wolbachia* strain *w*Au. BMC Genomics 15**:** 928.

Teixeira, L., A. Ferreira and M. Ashburner, 2008 The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. PLoS Biol. 6**:** e1000002.

Turelli, M., 1994 Evolution of incompatibility-inducing microbes and their hosts. Evolution 48.

Turelli, M., and A. A. Hoffmann, 1991 Rapid spread of an inherited incompatibility factor in California *Drosophila*. Nature 353**:** 440-442.

Turelli, M., and A. A. Hoffmann, 1995 Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations. Genetics 140**:** 1319-1338.

Turelli, M., J. R. Lipkowitz and Y. Brandvain, 2014 On the Coyne and Orr-igin of species: effects of intrinsic postzygotic isolation, ecological differentiation, X chromosome size, and sympatry on *Drosophila* speciation. Evolution 68**:** 1176-1187.

Turelli, M., B. S. Cooper, K. M. Richardson, P. S. Ginsberg, B. Peckenpaugh *et al.*, 2018 Rapid global spread of *w*Ri-like *Wolbachia* across multiple *Drosophila*. Curr. Biol. 28**:** 963-971.e968.

Turissini, D. A., and D. R. Matute, 2017 Fine scale mapping of genomic introgressions within 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.

1228 Weeks, A. R., M. Turelli, W. R. Harcombe, K. T. Reynolds and A. A. Hoffmann, 2007 From
1229   parasite to mutualist: rapid evolution of *Wolbachia* in natural populations of *Drosophila*.
1230   PLoS Biol. 5**:** e114.
1231 Weinert, L. A., E. V. Araujo-Jnr, M. Z. Ahmed and J. J. Welch, 2015 The incidence of bacterial
1232   endosymbionts in terrestrial arthropods. Proc. Biol. Sci. 282**:** 20150249.
1233 Werren, J. H., and D. M. Windsor, 2000 *Wolbachia* infection frequencies in insects: evidence of
1234   a global equilibrium? Proc. Biol. Sci. 267**:** 1277-1285.
1235 Werren, J. H., W. Zhang and L. R. Guo, 1995 Evolution and phylogeny of *Wolbachia*:
1236   reproductive parasites of arthropods. Proc. Biol. Sci. 261**:** 55-63.
1237 Wu, M., L. V. Sun, J. Vamathevan, M. Riegler, R. Deboy *et al.*, 2004 Phylogenomics of the
1238   reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by
1239   mobile genetic elements. PLoS Biol. 2**:** e69.
1240 Yen, J. H., and A. R. Barr, 1971 New hypothesis of the cause of cytoplasmic incompatibility in
1241   *Culex pipiens L*. Nature 232**:** 657-658.
1242 Yukilevich, R., 2012 Asymmetrical patterns of speciation uniquely support reinforcement in
1243   *Drosophila*. Evolution 66**:** 1430-1446.
1244 Zabalou, S., S. Charlat, A. Nirgianaki, D. Lachaise, H. Merçot *et al.*, 2004 Natural *Wolbachia*
1245   infections in the *Drosophila yakuba* species complex do not induce cytoplasmic
1246   incompatibility but fully rescue the *w*Ri modification. Genetics 167**:** 827-834.
1247 Zabalou, S., A. Apostolaki, S. Pattas, Z. Veneti, C. Paraskevopoulos *et al.*, 2008 Multiple rescue
1248   factors within a *Wolbachia* strain. Genetics 178**:** 2145-2160.

SUPPLEMENTAL INFORMATION

SUPPLEMENTAL MATERIALS AND METHODS

PCR amplification and Sanger sequencing of "A", "B", and "C" regions in Figure 5.

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

We evaluated regions A, B, and C in Figure 5 using PCR and Sanger sequencing. This includes one primer set covering the A region, four primer sets attempting to cover the B region, and one primer set covering the C region (Table S1). We performed touchdown PCR with an annealing gradient from 70°C to 55°C, which provided the same results as the standard PCR conditions. In cases where no products were returned, we lowered the annealing temperature from 55°C to 50°C and repeated the PCR. PCR products from a focal genotype (yakuba_S09_L24) were cleaned using the QiaQuick PCR purification kit (Qiagen, Germany). We sent the cleaned PCR products to Eurofins (Kentucky) for Sanger sequencing. The loci, primers, and product lengths of PCR experiments are presented in Table S1.

SUPPLEMENTAL TABLES

Table S1. Primers used in this study.

Locus/Target region	Primer name	Forward primer	Reverse primer	Reference
wsp: Wolbachia surface protein	*wsp*	5'TGGTCCAATAAGT GATGAAGAAAC-3'	5'AAAAATTAAACG CTACTCCA-3'	Braig *et. al.* 1998
2L Control: rDNA	*2L*	5'TGCAGCTATGGTC GTTGACA-3'	5'ACGAGACAATAA TATGTGGTGCTG-3'	Meany *et al.* 2019
YAK_*ISWpi1*, "A" in Figure 5[§]	Bridge_5A	5'TTTTTGTTCTCAG AGGTTGTCC-3'	5'GTTGCTGGAAAT GGTTGGAT -3'	Designed here
MEL_*ISWpi1*, "B" in Figure 5[*]	433808_2_694583	5'GCTTACCAAGGCA TGAATCC-3'	5'TGCTTATTTTTCT TGTGGTCCA-3'	Designed here
MEL_*ISWpi1*, "B" in Figure 5[*]	692607_2_433806	5'CCTCAGGTTCGAG TGACAATTT-3'	5'TAAGTTGTCTTCA ATATGCCGC-3'	Designed here
MEL_*ISWpi1*, "B" in Figure 5[*]	long_*ISWpi1*	5'TCCCACTAGCCTC TCTATTCCA-3'	5'ATACCGCCATTG ACACCCTG -3'	Designed here
MEL_*ISWpi1*, "B" in Figure 5[*]	long_*ISWpi1*_rev	5'CTGCTATTTCTCC CACTAGCCT-3'	5'GCACAAAGATAC GCCACCTG-3'	Designed here
YAK_CinB, "C" in Figure 5[†]	*cin*_region	5'CCGCTTGAGTTGG TACAGCTT -3'	5'TTATCAGAACCG CGCACAAG-3'	Designed here

[§]Produced a 960 bp product
[*]Failed to amplify
[†]Produced a 1,006 bp product

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

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

Table S4. Rate multipliers for the *Wolbachia* relaxed-clock chronograms. All numbers are estimated from the $\Gamma(2,2)$ 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}, $(9.6 \times 10^{-6}, 3.0 \times 10^{-5})$	1.2×10^{-5}, $(5.2 \times 10^{-6}, 1.9 \times 10^{-5})$	1.7×10^{-5}, $(8.5 \times 10^{-6}, 2.7 \times 10^{-5})$
*w*Mel	4.4×10^{-5}, $(2.2 \times 10^{-5}, 7.1 \times 10^{-5})$	5.7×10^{-5}, $(2.8 \times 10^{-5}, 9.2 \times 10^{-5})$	3.4×10^{-5}, $(1.5 \times 10^{-5}, 5.6 \times 10^{-5})$
D. yakuba- clade and *w*Mel *Wolbachia*	1.7×10^{-4}, $(1.1 \times 10^{-4}, 2.4 \times 10^{-4})$	1.4×10^{-4}, $(8.7 \times 10^{-5}, 1.9 \times 10^{-4})$	2.0×10^{-4}, $(1.2 \times 10^{-4}, 2.7 \times 10^{-4})$

Table S5. Divergence estimates and credible intervals for the *Wolbachia* relaxed-clock chronograms estimated from the $\Gamma(7,7)$ branch-rate prior.

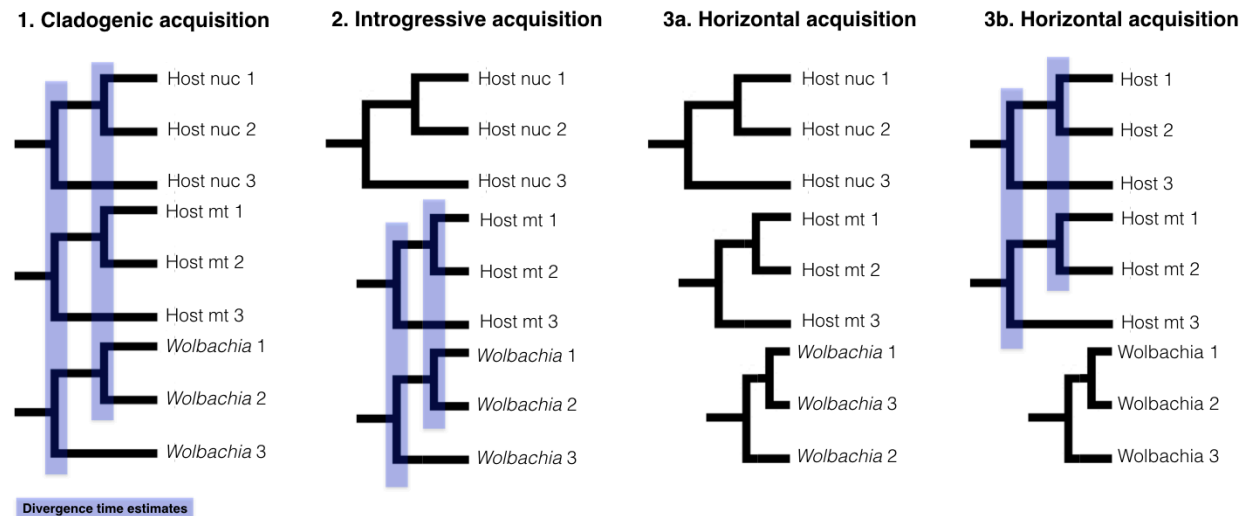| Relaxed-clock analysis	*w*Mel-*w*Yak divergence	*w*Yak-clade divergence	*w*Mel-clade divergence	*w*Yak-*w*San divergence
D. yakuba-Wolbachia	NA	2642, (781, 5904)	NA	1545, (401, 3680
*w*Mel	NA	NA	4955, (1363, 11093)	NA
D. yakuba- clade and *w*Mel *Wolbachia*	49585, (17452, 106310)	3914, (1432, 9024)	7908, (2511, 18739)	2252, (740, 5229)

SUPPLEMENTAL FIGURES

Figure S1 Phylogenetic hypotheses about *Wolbachia* acquisition. Infections could be 1) old and codiverging with host genomes (cladogenic acquisition), 2) codiverging with mtDNA (mt) with both mtDNA and *Wolbachia* having more recent divergence times than host nuclear genomes due to a history of host hybridization (introgressive acquisition), or 3) relatively young with more recent divergence times than host genomes (horizontal acquisition). For introgressive acquisition, mitochondrial and *Wolbachia* topologies could be discordant with the host nuclear topology. For horizontal acquisition, *Wolbachia* genomes could be discordant (3a) or concordant (3b) with host genomes. *Wolbachia* could be acquired horizontally by hosts with a history of hybridization introgression (3a) or by hosts with no history of hybridization and introgression (3b). Note that *Wolbachia* topologies could be concordant or discordant with host genomes in either case of horizontal acquisition.
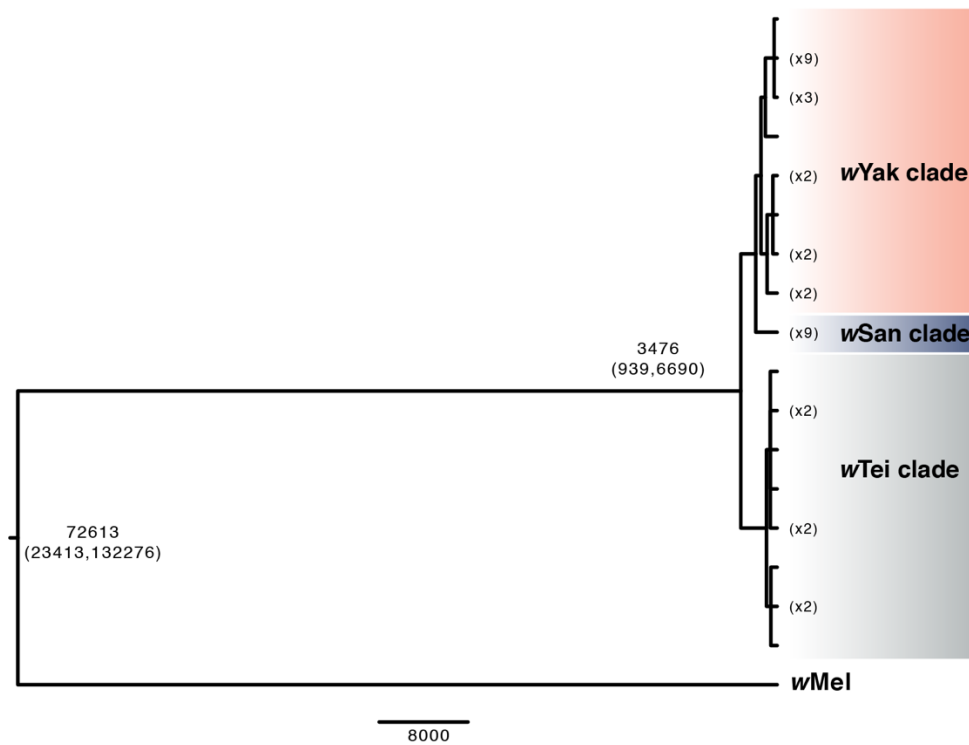
Figure S2 A Bayesian strict-clock *Wolbachia* absolute chronogram using 621 genes containing a total of 624,438 bp. Nodes with posterior probability less than 0.95 were collapsed into polytomies. Key nodes are labeled with age estimates.
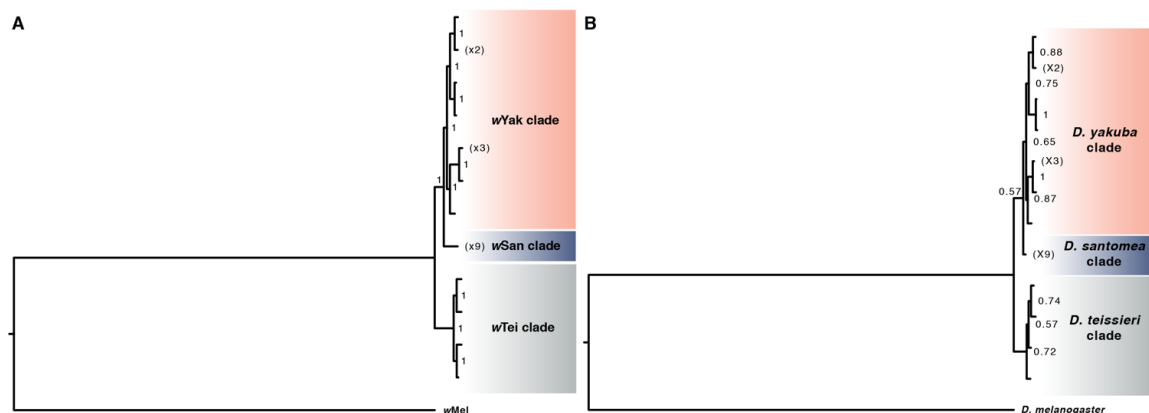

Figure S3 A. Bayesian *Wolbachia* phylogram using 621 genes containing a total of 624,438 bp. B. Bayesian mtDNA chronogram using the thirteen protein-coding genes. Nodes are labeled with their posterior probability.

Figure S4 A Bayesian phylogram with all branches fixed to length 1. The topology is the consensus of the mitochondrial and *Wolbachia* phylograms. Branches are labeled with the quartiles of the posterior distribution of ratios of the substitution rates of mtDNA versus *Wolbachia* along the branch (see Methods).