AEM Accepted Manuscript Posted Online 3 January 2020
Appl. Environ. Microbiol. doi:10.1128/AEM.02589-19
Copyright © 2020 American Society for Microbiology. All Rights Reserved.

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

1 ***Wolbachia*** **endosymbiont of the horn fly** ***Haematobia irritans irritans:*** **a supergroup A**
2 **strain with multiple horizontally acquired cytoplasmic incompatibility genes**

3

4 Mukund Madhav[a†], Rhys Parry[b†], Jess A.T. Morgan, Peter James[a], Sassan Asgari[b*]

5

6 † These authors contributed equally.
7 [a]Queensland Alliance for Agriculture and Food Innovation (QAAFI), The University of
8 Queensland, Brisbane, QLD 4072, Australia
9 [b]Australian Infectious Disease Research Centre, School of Biological Sciences, The
10 University of Queensland, Brisbane, QLD 4072, Australia

11

12 **Running title:** *Wolbachia* endosymbiont of the horn fly

13

14 **Word count** (abstract): 246
15 **Word count** (text): 5396
16 *Corresponding author: Sassan Asgari; Tel: +617 3365 2043; Fax: +617 3365 1655;
17 **s.asgari@uq.edu.au**

18

19 **Keywords:** horn fly, buffalo fly, Wolbachia, cytoplasmic incompatibility, phage

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

Accepted Manuscript Posted Online
Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology

20 **Abstract**
21 The horn fly, *Haematobia irritans irritans*, is a hematophagous parasite of livestock
22 distributed throughout Europe, Africa, Asia, and the Americas. Welfare losses on livestock
23 due to horn fly infestation are estimated to cost between USD 1-2.5 billion annually in North
24 America and Brazil. The endosymbiotic bacterium *Wolbachia pipientis* is a maternally
25 inherited manipulator of reproductive biology in arthropods and naturally infects laboratory
26 colonies of horn flies from Kerrville, USA and Alberta, Canada, but has also been identified
27 in wild-caught samples from Canada, USA, Mexico and Hungary. Re-assembly of PacBio
28 long-read and Illumina genomic DNA libraries from the Kerrville *H. i. irritans* genome project
29 allowed for a complete and circularised 1.3 Mb *Wolbachia* genome (*w*Irr). Annotation of *w*Irr
30 yielded 1249 coding genes, 34 tRNAs, three rRNAs, and five prophage regions.
31 Comparative genomics and whole genome Bayesian evolutionary analysis of *w*Irr compared
32 to published *Wolbachia* genomes suggests that *w*Irr is most closely related to and diverged
33 from *Wolbachia* supergroup A strains known to infect *Drosophila* spp. Whole-genome
34 synteny analyses between *w*Irr and closely related genomes indicates that *w*Irr has
35 undergone significant genome rearrangements while maintaining high nucleotide identity.
36 Comparative analysis of the cytoplasmic incompatibility (CI) genes of *w*Irr suggests two
37 phylogenetically distinct CI loci and acquisition of another *CifB* homolog from
38 phylogenetically distant supergroup A *Wolbachia* strains suggesting horizontal acquisition
39 of these loci. The *w*Irr genome provides a resource for future examination of the impact
40 *Wolbachia* may have in both biocontrol and potential insecticide resistance of horn flies.
41

42 **Importance**
43 Horn flies, *Haematobia irritans irritans*, are obligate hematophagous parasites of cattle
44 having significant effects on production and animal welfare. Control of horn flies mainly relies
45 on the use of insecticides, but issues with resistance have increased interest in development
46 of alternative means of control. *Wolbachia pipientis* is an endosymbiont bacterium known to
47 have a range of effects on host reproduction such as induction of cytoplasmic incompatibility,
48 feminization, male killing, and also impacts on vector transmission. These characteristics of
49 *Wolbachia* have been exploited in biological control approaches for a range of insect pests.
50 Here we report the assembly and annotation of the circular genome of the *Wolbachia* strain
51 of the Kerrville, USA horn fly (*w*Irr). Annotation of *w*Irr suggests its unique features including
52 the horizontal acquisition of additional transcriptionally active cytoplasmic incompatibility
53 loci. This study will provide the foundation for future *Wolbachia*-induced biological effect
54 studies for control of horn flies.

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology
Accepted Manuscript Posted Online

55 **Introduction**

56 Flies from the genus *Haematobia* (Diptera: Muscidae) are obligate hematophagous
57 ectoparasites of pastured cattle. Two prominent members of this genus are the horn fly,
58 *Haematobia irritans irritans*, distributed throughout Europe, Africa, Asia, and the Americas
59 (1) and the buffalo fly, *Haematobia irritans exigua*, which is widespread throughout Asia and
60 Australia (2). Blood-feeding behaviour from *H. i. irritans* results in severe welfare issues and
61 economic losses to cattle industries with annual estimates of up to $US ~1 billion in North
62 America and $US ~2.5 billion in Brazil (2-4). In Australia, *H. i. exigua* is estimated to cost
63 the domestic cattle industry $AUS 98.7 million annually and is currently restricted to the
64 northern part of the country (5). Control of *Haematobia* flies primarily relies on the use of
65 chemical insecticides; however, reports of insecticide resistance suggest that alternative
66 intervention strategies are required (2, 6, 7).

67 *Wolbachia pipientis* is an obligate, endosymbiotic, Gram-negative α-proteobacteria
68 estimated to infect between 40-70% of terrestrial arthropods (8, 9). *Wolbachia* infection in
69 insects is known to selfishly alter host reproductive biology to transmit and persist in the next
70 generation (10). One mechanism that drives transgenerational *Wolbachia* persistence is
71 known as cytoplasmic incompatibility (CI) (11, 12). In CI, mating between *Wolbachia*-
72 infected male and non-infected female (unidirectional CI) or female infected with a different
73 *Wolbachia* strain (bidirectional CI) results in embryo death (12). The commonly accepted
74 model for CI is “mod/resc”. Here, *mod* stands for modification of sperm by a toxin in the
75 *Wolbachia*-infected male, and *resc* for a rescue of sperm by an antidote present in the egg
76 (11, 13). Cellular studies have linked early embryonic death with defects in first zygotic
77 mitosis, irregular chromosomal condensation post-fertilisation, and delayed histone
78 deposition in the earlier interphase cell cycle (14-17). Two parallel studies recently identified
79 the molecular mechanisms underpinning CI. Using a combined genomic and transcriptomic
80 approach, LePage et al. (2017) identified two genes, *cifA* and *cifB*, in the prophage WO of
81 *w*Mel *Wolbachia* strain mediating CI (18). Whereas Beckmann et al. (2017) demonstrated
82 two genes *cidA* and cidB, *cifA* and *cifB* homologues, underpinned CI in the supergroup B
83 *Wolbachia* strain *w*Pip (19). Further experimental examination of the CI loci suggested a
84 “Two-by-One” model, whereby the *cifA* gene works as the rescue factor, and *cifA* and *cifB*
85 together instigated CI (20).

86 In addition to CI, other phenotypes of reproductive manipulation have been reported for
87 *Wolbachia* including male-killing, parthenogenesis, and feminisation (12). *Wolbachia* has
88 also been demonstrated to confer protection against RNA virus infection in dipteran hosts
89 (9, 10). Both CI and the ability of *Wolbachia* to restrict RNA viruses form the basis for the

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

Accepted Manuscript Posted Online
Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology

deployment of *Wolbachia*-infected *Aedes aegypti* mosquito for the control of dengue fever and other arboviruses worldwide (21, 22).

Wolbachia have been found to replicate in higher density in organophosphate resistant *Culex pipiens* mosquitoes than susceptible individuals (23, 24). However, no such association between insecticide resistance and *Wolbachia* density was observed in *Ae. aegypti* mosquitoes suggesting that interactions between the insecticide resistance phenotype and *Wolbachia* dynamics is both host and *Wolbachia* strain dependent (25).

While *H. irritans* are not currently known to be vectors of pathogenic viruses in livestock, there exists significant interest in exploiting the CI phenotype of *Wolbachia* as a form of sterile insect technique in *H. i. exigua* in Australia. A comprehensive screen of *H. i. exigua* samples from 12 locations in Australia and also Bali, Indonesia did not detect *Wolbachia* (26). By comparison *Wolbachia* has been previously identified in many wild-caught populations of *H. i. irritans* from Mexico (27), field-caught and laboratory colonies from the USA (28, 29), both field-collected and laboratory colonies from Alberta, Canada (26, 30), and also from field-collected samples in Hungary (31).

Due to the intracellular nature of *Wolbachia* and presence of multiple insertion sequences within *Wolbachia* genomes, assemblies using only short-read chemistries often result in highly fragmented assemblies (32). However, combining PacBio long-read sequencing and Illumina technologies has resulted in the closed and completed *Wolbachia* genome assembly (32, 33). The genome of the *H. i. irritans* Kerrville reference strain maintained at the USDA-ARS Knipling-Bushland U. S. Livestock Insects Research Laboratory (Kerrville, TX) was recently assembled using Pacific Biosciences (PacBio) SMRT technology and Illumina chemistries (34). Initial analysis of long and short read sequencing data indicated that a large portion of the reads in both libraries shared similarity to the *Wolbachia* endosymbiont of *Drosophila simulans* *w*Ri strain (34, 35). During the deposition of the *H. i. irritans* genome to NCBI *Wolbachia* contigs were removed (personal communication Felix Guerrero; USDA-lab, US). To recover the *Wolbachia* genome, we extracted the *Wolbachia* sequences from Kerrville *H. i. irritans* genome project and through bioinformatics analysis were able to assemble a high-quality and circularised *w*Irr genome. Further, we explored its phylogenetic relationship with the described *Wolbachia* strains, and the possibility of induction of CI by this strain based on what is known about the genes responsible for CI.

Results and Discussion

*w*Irr genome assembly, annotation and genome features

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

125 To extract and assemble the genome of *Wolbachia* from *H. i. irritans*, the genomic data from
126 the Kerrville reference genome project (34) was trimmed and mapped against published
127 *Wolbachia* genomes (32, 35-37) using Burrows-Wheeler Aligner (BWA-MEM) under relaxed
128 mapping criteria (38). Initially, ~10 million of ~404 million paired-end Illumina reads and
129 128,203 of 4,471,713 (2.86%) PacBio reads mapped to representative supergroup A
130 *Wolbachia* genomes. These reads were then extracted, and *de novo* assembled using
131 Unicycler resulting in a singular, circularised draft assembly (39). Raw Illumina fastq reads
132 were then iteratively mapped against this draft genome and polished using pilon (40). The
133 final number of reads that mapped to the assembled *Wolbachia* genome were 140,429 out
134 of 4,471,713 (3.1%) from the PacBio library, corresponding to an average coverage of
135 ~187x, and 10,285,275 out of 404,202,898 (2.54%) from the paired-end Illumina libraries,
136 corresponding to an average coverage of ~1280x. The final *w*Irr genome is 1,352,354bp
137 with a GC content of 35.3%, which is similar to other previously assembled supergroup A
138 *Wolbachia* strains (Table 1). The polished *w*Irr genome was then annotated using the NCBI
139 prokaryotic genome annotation pipeline (41) which predicted that *w*Irr encodes for 1,419
140 genes with 1,249 protein-coding genes and 129 pseudogenes, with 56 containing
141 frameshifts, 93 incomplete, 12 with an internal stop, and 31 with multiple problems. The RNA
142 gene repertoire of the *w*Irr genome was identified to encode 34 tRNAs, three rRNAs (5s,
143 16s, and 23s), and also non-coding RNA genes such as RNase P RNA component class A
144 (RFAM: RF00010), signal recognition particle sRNA small type (RFAM: RF00169), 6S RNA
145 (RFAM: RF00013) and transfer-messenger RNA (RFAM: RF01849). Completeness of the
146 *w*Irr genome was assessed by comparing the proteome against 221 single-copy orthologs
147 derived from 1520 proteobacterial species in BUSCO pipeline (42). The BUSCO score for
148 completeness of a model organism with a good reference genome is usually above 95%,
149 but for the endosymbiotic bacteria with degenerated metabolic pathways BUSCO scores
150 can vary between 50% to 95% based on the genome size, presence of repetitive elements
151 in the genome and individual taxonomic placement (43). The completeness score for *w*Irr
152 was 82.4%, which included 182 single-copy orthologs, two fragmented and 37 missing
153 orthologs (Fig. S1), similar to five other completed *Wolbachia* genome projects (*w*Au, *w*Mel,
154 *w*Ha, and *w*Ri).
155 We compared the proteome of *w*Irr against the four completed and circularised supergroup
156 A *Wolbachia* strains (*w*Au, *w*Mel, *w*Ha, and *w*Ri) using the Orthovenn 2 web server (44). A
157 total of 1136 conserved orthologs were identified in all five strains. All five strains shared
158 810 orthologs with 782 of these being single-copy (Fig. S2). The *w*Irr genome has 1005
159 orthologs comprising of 1248 proteins, mostly involved in cell function and metabolism.

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

160 While seven clusters of "singletons" were predicted to only exist in *w*Irr by Orthovenn, closer
161 examination of six of these clusters suggests these are transposable elements that are
162 present but not annotated in the Genbank *Wolbachia* genome assemblies. These will be
163 explored further below.
164 In addition to DNA sequencing data, we explored the transcriptional activity of *w*Irr in all life
165 stages of *H. i. irritans* by mapping RNA-Seq data used to annotate the genome. As each
166 sample was only sequenced once and poly-A enriched, it is difficult to make differential gene
167 expression analyses with the data or infer *Wolbachia* tissue distributions. However, it
168 appears *w*Irr is present and transcriptionally active in all life stages and all tissues dissected
169 (Table 2). There is lower transcriptional activity in eggs and pupae than adults and also the
170 highest normalised transcriptional activity was found in adult libraries at two hours post blood
171 meal.
172

173 **Phylogenetic placement of *w*Irr suggests close relationship with *Drosophila* spp.**
174 **supergroup A *Wolbachia* strains**

175 Since the discovery of *Wolbachia* within the gonads of the *Culex pipiens* mosquito,
176 *Wolbachia* has taxonomically been considered a single species divided into 16 major
177 supergroups (denoted A-Q) (45, 46). While the suitability of classifying the supergroups into
178 a single *Wolbachia* species is the subject of ongoing debate (47, 48), a universal genotyping
179 tool has been developed to demarcate supergroups based on multilocus sequence typing
180 (MLST) of five ubiquitous genes (*gatB*, *coxA*, *hcpA*, *fbpA*, and *ftsZ*) (49). Although MLST
181 clearly demarcates *Wolbachia* strains to supergroups, it fails to reliably discriminate strains
182 within supergroups with high phylogenetic support. As such, a recent examination of these
183 loci by Bleidorn and Gerth (2018) suggests that a number of alternative single copy loci
184 outperform these five genes (49, 50). To construct a whole-genome phylogenetic analysis
185 of *w*Irr, we used 79 of the 252 single copy orthologs from non-recombinant loci identified by
186 Bleidorn and Gerth (2018) from 19 strains of *Wolbachia* (50). The phylogeny gives strong
187 posterior probability support for the *Wolbachia* *w*Irr strain being basal to a clade containing
188 *w*Rec, *w*Au and *w*Mel in supergroup A (Fig. 1).
189 Natural *Wolbachia* transfer between hosts can be cladogenic (*Wolbachia* acquired during
190 the speciation of hosts), introgressive (transfer during mating between closely related host
191 species), or horizontal (possibly via shared food and ecological niche, wounds and vectors)
192 (51, 52). Concordance between the *Wolbachia* genome with the hosts mitochondrial and
193 nuclear genome with consistent divergence time shows cladogenic transfer, whereas
194 discordance suggests the possibility of horizontal transmission. Taxonomically, all

Accepted Manuscript Posted Online
Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology

195 Drosophilidae belong to the Ephydroidea superfamily of muscomorph flies, in which the
196 *Wolbachia* strains *w*Au, *w*Ri, *w*Mel and *w*Rec have been identified. The *Haematobia* genus
197 belongs to the superfamily Muscoidea in the insect order Diptera (53). Divergence estimates
198 of Ephydroidea and Muscoidea inferred from mitochondrial genes suggests that the most
199 recent common ancestor of all *Haematobia* and *Drosophila* diverged sometime in the
200 Palaeocene ~60 Million years ago (Mya) (54).
201 A number of phylodynamic analyses of *Wolbachia* genomes have attempted to reconstruct
202 evolutionary timescales, albeit with limited concordance between analyses (55-57). One
203 Bayesian time to most recent common ancestor (TMRCA) analysis conducted by Gerth and
204 Bleidorn (2016) on the clade encompassing all *Drosophila Wolbachia* strains was dated at
205 48.38 Mya, with a range of 110 – 16 Mya. This fits within the mitochondrial divergence of
206 *Haematobia* from *Drosophila* (~60 Mya) and divergence of *Wolbachia* from supergroup A
207 members (*w*Mel, *w*Ri, and *w*Rec). Due to the limited genetic data available for *Wolbachia*
208 infecting *Haematobia* we did not make an attempt to formally test the mode of transmission
209 through timescale estimates or phylogenetic discordance as we may potentially incorrectly
210 conclude the mode of transmission. As such, it still remains to be elucidated if the close
211 genetic relationship between *w*Irr and other *Wolbachia* may be the result of codivergence or
212 horizontally acquired in *H. i. irritans.*
213
214 **The Kerrville *Wolbachia w*Irr strain is closely related to wild *H. i. irritans Wolbachia***
215 **strains from the US, Mexico, Canada and Hungary**
216 Previous publications have demonstrated the presence of *Wolbachia* from wild-caught and
217 laboratory colonies of *H. i. irritans* through amplicon Sanger sequencing of samples (29, 31,
218 58) or identifying *Wolbachia* reads in pyrosequencing-based approaches or expressed
219 sequence tags (EST) (27, 28). Currently, avaliable Genbank data from *Wolbachia* of *H. i.*
220 *irritans* are limited to partial fragments of the *Wolbachia surface protein* (*wsp*) gene (29, 58)
221 or fragments of the *16S ribosomal RNA* gene (31). BLASTn analysis of the *wsp* fragment
222 sample of the Kerrville colony used by Jeyaprakash and Hoy (2000), designated as wIrr-A1
223 (Genbank: AF217714.1) (29), showed 100% identity with the *wsp* locus of *w*Irr (Gene:
224 E0495; Position: 1,282,799-1,283,488). Similar high nucleotide identity of the *wsp* fragment
225 of *H. i. irritans* samples, originating from Lethbridge, Alberta, Canada designated *w*Irr
226 (Genbank: DQ380856.1), with the *w*Irr *wsp* was found; 99.64% with only two nucleotide
227 differences over an amplicon of 554bp. In addition, the *w*Irr *16s rRNA* gene (Position:
228 882,502-884,006) and partial 16S rRNA fragments from two *Wolbachia* strains from *H. i.*
229 *irritans* Hungary samples (Genbank: EU315781.1, EU315780.1) were 99.62% identical with

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

Accepted Manuscript Posted Online

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

230 264/265 sequence identity. While this suggests the *w*Irr strain of *Wolbachia* is very closely
231 related to the Canadian and Hungarian *H. i. irritans* samples, the nature of the amplicon size
232 and the high nucleotide identity between strains make it difficult to state this with complete
233 certainty.
234 As high-throughput sequencing allows for a closer examination of relatedness between the
235 Kerrville *w*Irr *Wolbachia* strain and wild-caught *H. i. irritans* harbouring *Wolbachia*, we re-
236 analysed EST, DNA-Seq and RNA-Seq data with BLASTn using our *w*Irr genome as a query
237 from a number of publications using wild-caught flies from Mexico, USA and also Uruguay
238 (Table 4). We identified five EST fragments, and 394 assembled *Wolbachia* RNA contigs
239 from wild-caught *H. i. irritans* from two different studies of Louisiana State University
240 Agricultural Center St. Gabriel Research Station (LA, USA) (59, 60), and four EST fragments
241 from a cattle farm in Ciudad Victoria, Tamaulipas, Mexico (61). Additionally, in six RNA-Seq
242 libraries of newly emerged male and female horn flies wild-caught in Louisiana, USA on
243 average 10% of each library could be mapped to the *w*Irr genome (Table S1) (59). All
244 identified contigs shared closer nucleotide identity to the *w*Irr strain than any other
245 *Wolbachia* genome deposited on NCBI (data not shown). Interestingly, we could not identify
246 any assembled contigs or reads that mapped to the *w*Irr genome from salivary gland and
247 midgut samples originating from wild-collected *H. i. irritans* from Canelones, Uruguay (62)
248 suggesting that either *Wolbachia* is present in very low abundance in these samples or
249 completely absent.
250 We conducted *de novo* assembly of the 454 DNA-Seq data originating from a single male
251 *H. i. irritans* collected in 2003 from the Pressler Cattle Ranch in Kerrville, Texas, USA (63).
252 Of 1,130 assembled contigs 74 were identified through BLASTn analysis as having closest
253 bit score hit to the *w*Irr genome (data not shown). As very few *Wolbachia* genome fragments
254 were conserved from RNA-Seq and DNA-Seq assemblies, we could not construct a single
255 phylogenetic tree for all the samples. However, the close identity of all available
256 transcriptome and genomic data of wild-caught *H. i. irritans* flies from North American
257 populations including Mexican to the *Wolbachia* Kerrville reference *H. i. irritans* strain
258 suggest that likely they are also infected with the same *w*Irr strain.
259

260 **The *w*Irr genome has undergone significant genome rearrangements compared to**
261 **other *Wolbachia* genomes**

262 In bacterial genome evolution, horizontal gene transfer (64, 65) and genetic vehicles such
263 as bacteriophages, plasmids or transposons (mobile element) (65-68) contribute to changes
264 in the bacterial genome. Due to the intracellular niche of the endosymbiont, the evolution of

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

265 *Wolbachia* genomes is highly dependent on bacteriophages, and transposable elements,
266 with both contributing to sometimes as much as 21% of the genome (65). Whole-genome
267 comparisons of nucleotide synteny between *w*Irr and *w*Mel and *w*Ri were carried out using
268 MAFFT v.7 (69). We did not analyse the synteny between *w*Rec (70) and wIrr because the
269 genome is fragmented and yet to be circularised. It appears that while *w*Irr maintains
270 between 90-99% nucleotide identity with the other two strains, *w*Irr has undergone a high
271 degree of genome rearrangement (Fig. 2A and B). In comparison, *w*Mel and *w*Ri show very
272 similar genome arrangements (Fig. 2C). Similar genomic rearrangement has been
273 previously seen while comparing *w*Pip and *w*Mel, *w*Mel and *w*Bm, *w*Uni and *wVitA* (71-73).
274

275 **Expansion of insertion sequence elements in *w*Irr genome is associated with a**
276 **divergent CifB homologue**

277 Insertion sequences (IS) are diverse transposable elements in bacterial genomes (65, 74).
278 Considerable variation in the IS element composition in *Wolbachia* genomes is speculated
279 to contribute to diversification or speciation of closely related strains, and IS elements can
280 cause the disruption of protein coding genes leading to pseudogenes (32, 36). To compare
281 the IS element load between *w*Irr and other supergroup A *Wolbachia*, *w*Ri, *w*Au, *w*Mel, *w*Ha,
282 IS elements were identified and searched against the IS finder database using the ISsaga
283 web server (74) (Supplementary File 1). A total of 283 ORFs related to IS elements were
284 identified in the *w*Irr genome, including 61 complete ORFs and 150 partial IS elements.
285 Maximum copies of IS elements were from IS630 (111 copies), which belong to the
286 Tc1/mariner (Class II) transposon family, and ssgr IS1031 (109 copies), which is from the
287 IS5 family. Comparative analyses between *w*Irr and other supergroup A *Wolbachia* strains
288 identified 12 conserved IS families between all genomes IS66_ssgr_ISBst12, ISL3,
289 IS5_ssgr_IS1031, IS4_ssgr_IS4, IS4_ssgr_IS231, IS3_ssgr_IS3, IS110,
290 IS110_ssgr_IS1111, IS4_ssgr_IS50, IS630, IS481 and IS5_ssgr_IS903. However, two IS
291 families were identified as exclusive to *w*Irr: IS5_ssgr_IS427, which has one complete ORF
292 and three partial ORFs, and the IS5_ssgr_ISL2, with two partial ORFs. We manually
293 extracted the IS5_ssgr_IS427 annotations and within one of the identified loci (positions
294 632,890 and 630,128), a disrupted IS5-like element was found with only one single hit.
295 Based on BLASTn similarity (Query length:100%, Nucleotide identity: 80.39% E-value: 0),
296 this element is from the *Wolbachia* endosymbiont of *Brugia malayi* isolate TRS (Genbank
297 ID: CP034333.1) (72). Immediately after this transposable fragment is the protein
298 E0495_03245 (Fig. 3A), which BLASTp analysis of this 546aa protein appears to be a
299 truncated CI factor *cifB* belonging to the *w*Ha *Wolbachia* endosymbiont of *Drosophila*

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

300 *simulans* (Genbank ID: WP_144054595.1, Query cover: 98% Percentage similarity: 65.71%
301 E-value: 0.0). We examined the transcriptional activity of this *cifB* gene by mapping the RNA-
302 Seq data of all life stages to this region of the genome. One paired read mapped to this gene
303 at this location suggesting reduced transcriptional activity compared to other cifA/cifB loci
304 (Fig. 3B). However, due to the nature of the RNA-Seq library preparation with polyA
305 enrichment, we are unable to make strong conclusions and comparisons between different
306 genomic loci. The length of IS elements varied between 174 to 1743 bp having a median
307 size of 348bp. The total burden of IS elements on the *w*Irr genome is 115,692 bp, which is
308 8.55%. This is similar to the IS element percentage found in *w*Ri (9%) which is double that
309 of the IS element load of *w*Mel (4.3%), *w*Ha (4.4%), and *w*Au (4.4%). This lineage-specific
310 attainment and loss of IS elements, as well as length of the IS element, size and family
311 distribution is well documented across *Wolbachia* strains (36). The discovery of a single IS
312 element shared between *w*Irr (Supergroup A) and *Wolbachia* from the filarial nematode
313 *Brugia malayi* (Supergroup D) is of particular interest. *B. malayi* is a filarial nematode that
314 relies on a hematophagous mosquito host as a vector. Potentially, the gain of this IS element
315 may have arisen through co-infection of *H. i. irritans* with a distantly related nematode
316 species as it seems unlikely to have been independently lost in all other supergroup A
317 genomes. While *H. i. irritans* is known to vector *Stephanofilaria* sp. nematodes (75),
318 presence or absence of *Wolbachia* within these nematodes is yet to be characterised, and
319 therefore formal testing of IS acquisition cannot be undertaken. Further assembly and
320 genetic characterisation of filarial nematodes and their *Wolbachia* endosymbionts would
321 allow for a better understanding of interaction between the *H. i. irritans*, *Stepahnofilaria* sp.
322 and *Wolbachia.*
323
324 **The prophage regions of *w*Irr have a reduced eukaryotic association module**
325 *Wolbachia* bacteriophages or prophages (WO) have been widely reported in strains from
326 supergroup A, B and F, however, they have been lost in supergroup C and D strains (76).
327 The tripartite relationship between *Wolbachia*–WO and arthropod hosts is of great interest
328 as it has been shown that many genes located within prophage regions of *Wolbachia*
329 genomes contain eukaryotic association genes and toxin-antitoxin modules (77), and also
330 there is interest in utilising WO as a candidate for *Wolbachia* genetic transformation (76, 78).
331 Using the Phaster web server, we identified five potential WO regions in the *w*Irr genome.
332 The largest of which is a 60.8kb region designated as "intact" by Phaster with 68 ORFs from
333 359,527-420,415 having head, baseplate, tail, virulence genes and IS630 family
334 transposons (79). The other four were ~7Kb incomplete prophage regions containing 10, 9,

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

335 12 and 8 ORFs positioned at 613,245-620,397, 859,203-866,672, 903,423-910,665, and
336 1,241,523-1,247,571 respectively in the *w*Irr genome. Supergroup A members *w*Mel, *w*Ri,
337 *w*Au, and *w*Ha have between two to four variable WO phage regions with at least one
338 presumed intact and other WO-like degenerated phage regions (32, 35-37). We compared
339 the "intact" putative prophage region of WOIrr with the predicted WO phage regions from
340 *w*Mel (WOMelB) and completely sequenced WO phage region from *w*VitA (WOVitA), to
341 identify the conserved region using reciprocal BLASTn analysis (37, 65, 80). The conserved
342 phage regions were visualised using Easyfig (Fig. 4). Previously, it has been reported that
343 a eukaryotic association module (EAM) is present in the WO phage genomes from
344 *Wolbachia* characterised by proteins that are enriched with eukaryotic-like domains (77).
345 EAM are enriched for ankyrin repeat domains (ANK), which are involved in regulation of cell
346 cycle, promotion of protein-protein interactions, and *Wolbachia*-induced reproductive
347 phenotypes (81-83) and vary widely between strains (84). Reciprocal blast comparisons
348 between WOVitA and WOIrr (Fig. 4A) suggest there is a reduction of the EAM in WOIrr. In
349 WOVitA, the EAM lies between the hypothetical protein *gwv_1089* and the patatin-like
350 phospholipase family protein *gwv_1104*. In WOIrr, there is no ankyrin repeat with PRANC
351 domain protein (*gwv_1092*), ankyrin and tetratricopeptide repeat family protein (gwv_1093).
352 These are also not found in chromosomal *w*Irr. Additionally, the only protein that is
353 conserved between the EAM module is the patatin-like phospholipase *E0495_02120* which
354 shares 78.88% pairwise protein identity with BLASTp analysis with *gwv_1104*. The EAM is
355 also missing in WOMelB (Fig. 4B).
356
357 **Horizontal acquisition of *Wolbachia* cytoplasmic incompatibility loci in *w*Irr**
358 To explore the genetic diversity of CI genes in *w*Irr, we explored orthologous clusters for the
359 previously described CI genes. In addition to the truncated *cifB* (E0495_03245) gene, we
360 found two complete and genetically distant CI operons in *w*Irr, with one located within the
361 WOIrr region (Gene ID: E0495_02160, E0495_02165) and the second (Gene ID:
362 E0495_02270, E0495_02275) downstream of the WOIrr region. BLASTp analysis of the
363 predicted protein sequences (Table 3) indicated that these CI genes are not duplications as
364 previously reported for *w*Ri (35).
365 The CI genes of *Wolbachia* have been grouped into four different phylogenetic groups (Type
366 I - IV) (18, 85), as such, we conducted a Bayesian phylogenetic analysis of the complete CI
367 genes of *w*Irr (Fig. 4). For one set of CI genes (Genes: E0495_02270, E0495_02275)
368 located just outside the predicted WO region of *w*Irr, both copies of *cifA* and *cifB* genes
369 phylogenetically clustered to a clade encompassing *w*Ri and *w*Mel (Fig. 5) showing high

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

370 posterior probability support. The second *cifA/B* gene set, which sits within the WOIrr region
371 (Genes: E0495_02160, E0495_02165), groups phylogenetically with *Wolbachia cifA/B*
372 genes originating from WOSol, which is a WO phage infecting the *Wolbachia* strain (*w*Sol)
373 from the fig wasps *Ceratosolen solmsi* (86), and *w*Ha, which is a *Wolbachia* strain infecting
374 *Drosophila simulans* (Fig. 1) that is from a *Wolbachia* clade evolutionarily distinct from *w*Irr.
375 Based on phylogenetic discordance and high posterior support for ancestors of *CifA/B*
376 genes, there is reasonable support for the horizontal acquisition of E0495_02160, and
377 E0495_02165 genes in the *w*Irr genome from other distantly related *Wolbachia*. This report
378 is similar to another independent acquisition of CI genes in the *Wolbachia* endosymbionts
379 of the *Drosophila yakuba* clade which cause weak intra-and interspecific CI (87). Cooper et
380 al. (2019) assembled the genomes of *w*Yak variants and demonstrated that while there
381 appears to be another CI locus in these genomes, the presence of an inversion introduces
382 several stop codons within the cidB$^{wYak\text{-}clade}$ locus relative to the same region in cidB^{wMel},
383 speculated to potentially render this gene non-functional (88). To support the horizontal
384 acquisition of these CI loci in *w*Yak, Cooper and collegues compared pairwise differences
385 between the homologues and nuclear genes. By comparison, both genes within the CI loci
386 in *w*Irr are seemingly complete with no premature stops and presumed to encode for
387 functional proteins. Previous studies have suggested that the CI gene sets *cifA* and *cifB* vary
388 in copy number across CI-inducing *Wolbachia* strains and are directly correlated with the
389 extent of CI (strong or weak) (85). The acquisition of a second set of CI genes corroborates
390 previously unpublished experiments conducted where *w*Irr *Wolbachia* from the Kerrville
391 reference strain demonstrated a strong CI phenotype (personal communication Felix
392 Guerrero; USDA-lab, US). The transcriptional activity of the CI genes have previously been
393 explored by Lindsey et al. (2018) who demonstrated that both *cifA* and *cifB* show differential
394 transcriptional activity across host development (85). Again, RNA-Seq data of all life stages
395 were mapped to the *w*Irr genome and we examined the mapped reads at these two CI loci.
396 Reads mapped exclusively to one CI region and very few reads mapped to both (MAPQ
397 score 0). In general, the *cifA* gene was more transcriptionally active than the *cifB* gene in
398 both loci, as also previously reported (Fig. 6) (85). The evidence of two transcriptionally
399 active CI loci may explain the high incidence of *Wolbachia* in wild-caught specimens of *H. i.*
400 *irritans* as *Wolbachia* has been identified in 100% of all collected individuals from Hungry
401 (10/10) (31), as well as all 15 tested horn flies from two wild locations in Alberta, Canada
402 and also in 54/55 individuals tested in two independent screens of the laboratory colony of
403 Lethbridge Research Centre, Alberta, Canada (26).

404 **Conclusion**

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

405 In this study, we assembled and annotated a high-quality genome of the *Wolbachia*
406 endosymbiont of *H. i. irritans* designated *w*Irr. Phylogenetic analysis of the *w*Irr strain
407 suggests that the *w*Irr belongs to a well-supported supergroup A lineage that includes the
408 well-studied *w*Mel, *w*Au, and *w*Ri *Wolbachia* strains from *Drosophila* spp. Comparative
409 genomics of *w*Irr indicated acquisition of additional transcriptionally active CI loci.
410 Phylogenetic analysis indicates either horizontal acquisition of these genes from a closely
411 related *Wolbachia* strain or the potential loss of CI loci in other *Wolbachia* strains infecting
412 *Drosophila spp*. The *w*Irr genome has undergone significant reassortment compared to
413 closely related and completely assembled strains. Additional analysis of available and
414 deposited sequencing data from wild-caught and laboratory *H. i. irritans* colonies suggest
415 that *w*Irr is the most closely related to wild USA and Mexican samples and close relative of
416 Canadian and Hungarian samples. This study provides the foundation for future functional
417 studies of effects *Wolbachia* may have on life-history traits of *H. i. irritans* such as insecticide
418 resistance and evaluating contribution of *w*Irr towards population control.
419
420 **Materials and Methods**
421
422 **Genomic DNA and RNA Sequencing data**
423 The Kerrville reference *H. i. irritans* strain is a closed fly colony which has been maintained
424 at the USDA-ARS Knipling-Bushland U. S. Livestock Insects Research Laboratory since
425 1961 (34). Genomic DNA from unfed adult flies of mixed-sex originating from this strain was
426 subjected to whole-genome sequencing, and previously deposited on the National Center
427 for Biotechnology Information Short Read Archive (SRA) (Accession number: PRJNA30967)
428 (34). Briefly, this data includes two PacBio runs; one 10 kb and two 20 kb insert libraries. 10
429 kb libraries were sequenced using C2 chemistry and P4 polymerase, whereas C3 chemistry,
430 and P5 polymerase were used for both 20kb libraries with 3 hours of movie time. 10 kb
431 libraries and two of the 20 kb libraries were sequenced on 12 SMRTCells, four SMRTCells,
432 and eight SMRTCells, respectively, and all the sequences were finally pooled and uploaded
433 under the same accession (SRA: SRR6231657). For Illumina sequencing, one short-insert
434 paired-end library and one mate-paired end library with 6-12 kb insert size were sequenced
435 as 100nt paired ends on the HiSeq2000 and uploaded under the same accession number
436 (SRA: SRR6231656). Additional RNA sequencing data from different life stages and tissues
437 of the horn flies were sequenced on a Illumina HiSeq 2000 using 2x 100nt configuration and
438 available with the above Illumina read accession number (SRA: SRR6231656).
439

Accepted Manuscript Posted Online

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

440 ***Wolbachia* genome assembly, polishing and RNA-Seq analysis**

441 Raw fastq files originating from Illumina and PacBio sequencing data were imported to the
442 Galaxy Australia webserver (https://usegalaxy.org.au/, version 19.05; accessed between
443 May to October 2019). The Nextra universal transpose Illumina sequence adapters were
444 removed and reads were quality trimmed using Trimmomatic (Galaxy version: 0.36.4) under
445 the following conditions (Sliding window=4, average quality=20) (89). Resultant clean reads
446 were mapped to the genome of *w*Ri (35), and *w*Au (32) using BWA-MEM (Galaxy Version
447 0.7.17.1) (38) under default parameters and under simple Illumina mode and PacBio mode
448 (-x pacbio) for subsequent libraries. Mapped reads were extracted using a BAM filter (Galaxy
449 Version 0.5.9) and were then assembled using Unicycler (Galaxy Version 0.4.1.1) (39). For
450 RNA-Seq analysis, we also used BWA-MEM (Galaxy Version 0.7.17.1) (38) under default
451 parameters and under simple Illumina mode. To visualize mapped RNA-Seq data, resultant
452 BAM files were visualized with Integrated Genomics Viewer (IGV v 2.5.2).

453

454 **Genome annotation and comparative genomics**

455 Coding regions and ncRNAs of the assembled *w*Irr genome contig were annotated using
456 the NCBI prokaryotic genome annotation pipeline (41). To assess the quality of the
457 assembly, BUSCO v. 3.1.0 was used to search for orthologs of the near-universal, single-
458 copy genes in the BUSCO proteobacteria database (42). As a control, we performed the
459 same search using the reference genomes for *w*Ri (35), *w*Au (32), *w*Mel (37), *w*Ha,
460 and *w*No (36) as well as the complete *w*AlbB genome (33). Identification of phage and
461 prophage regions of *w*Irr was conducted using the PHASTER web platform
462 (https://phaster.ca/; accessed September 4, 2019) (79). Groupings of orthologous clusters
463 were identified using the Orthovenn2 web server (https://orthovenn2.bioinfotoolkits.net/;
464 accessed May 5, 2019) (44) under the following conditions: E-value: 1e-2, Inflation value:
465 1.5. Insertion sequence (IS) elements of *w*Irr were identified using the ISsaga web server
466 platform (http://issaga.biotoul.fr/; accessed August 8, 2019) (90). For nucleotide synteny
467 plots of *w*Irr MAFFT (https://mafft.cbrc.jp/alignment/server/; accessed July 8, 2019) (91) was
468 used to align *w*Irr and other genomes and then visualised by dot-plots of matches (without
469 extensions) identified using the LAST algorithm which compares sequences by adaptive
470 and fixed-length seeds (score=39, E=8.4e-11). Comparisons between the putative
471 prophage regions of *w*Irr were examined using BLASTn and visualised using Easyfig v. 2.2.2
472 (92).

473

474 **Phylogenetic analyses of *w*Irr and cytoplasmic incompatibility loci**

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

475 For full genome phylogenetic analyses, we used 79 non-recombinant gene loci, which has
476 been previously determined by Bleidorn & Gerth (2018) to perform well from 19 strains of
477 *Wolbachia* (50). These were downloaded (https://github.com/gerthmicha/wolbachia-mlst;
478 accessed September 2019), aligned using MUltiple Sequence Comparison by Log-
479 Expectation (MUSCLE v3.8.98) installed as part of the CLC Genomics Workbench (Version
480 11.0.1) (93) and concatenated. The resultant alignment was analysed using Bayesian
481 evolutionary analysis by sampling trees (BEAST v2.5.1) (94), split into individual codon
482 positions with linked site model, and unlinked clock model under the General Time
483 Reversible and Gamma = 4 nucleotide substitution model. Clock rates were drawn from a
484 log-normal distribution. Additional parameters were a chain length of 10 million steps
485 sampling every 10,000 steps under a Yule model. For phylogenetic placement of the CI
486 genes within *w*Irr, identified Cif homologs were first aligned using MUSCLE (93) and also
487 subjected to BEAST (94) with 10 million steps with a pre-burnin of 100,000 with sampling
488 being conducted every 20,000 steps under a Yule model and a general empirical model of
489 protein evolution (WAG) amino acid substitution model. For both BEAST runs convergence
490 for all parameters as well as stationary distributions of the MCMC chain were inspected
491 using Tracer v1.7.1 (effective sample sizes of >400). The maximum clade credibility (MCC)
492 tree (i.e. the tree with the largest product of posterior clade probabilities) was selected from
493 the posterior tree distribution using the program TreeAnnotator (included in the BEAST
494 package) after a 10% burn in. Resultant MCC trees were then visualised using FigTree
495 v1.4.4.
496
497 **Data availability and accession numbers**
498 PacBio and Illumina raw sequencing data are available from the NCBI short read archive
499 under accession numbers SRR6231657 and SRR6231656, respectively. The assembled
500 *Wolbachia pipientis w*Irr strain has been deposited in Genbank under the accession number
501 CP037426. Additional sequencing data and metadata used for validation are available in
502 supplementary files.
503
504 **Acknowledgements**
505 The authors acknowledge the support of Felix Guerrero from the USDA-ARS Knipling-
506 Bushland US Livestock Insects Research Laboratory and the suggestions and
507 reccomendations by two anonymous reviewers. Analysis was conducted using the
508 Australian Galaxy platform (https://usegalaxy.org.au/) with the support and technical

Applied and Environmental Microbiology AEM Applied and Environmental Microbiology Accepted Manuscript Posted Online

509 assistance of Igor Makunin. This project was funded by the Australian Research Council
510 grant (DP150101782) to SA and the University of Queensland scholarship to M.M. and R.P.

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

Accepted Manuscript Posted Online
Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology
Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

511 **References**

512 1. Changbunjong T, Weluwanarak T, Samung Y, Ruangsittichai J. 2016. Molecular
513 identification and genetic variation of hematophagous flies, (Diptera: Muscidae:
514 Stomoxyinae) in Thailand based on cox1 barcodes. J Asia-Pacific Entomol 19:1117-
515 1123.
516 2. Oyarzún M, Quiroz A, Birkett M. 2008. Insecticide resistance in the horn fly:
517 alternative control strategies. Med Vet Entomol 22:188-202.
518 3. Cupp E, Cupp M, Ribeiro J, Kunz S. 1998. Blood-feeding strategy of *Haematobia*
519 *irritans* (Diptera: Muscidae). J Med Entomol 35:591-595.
520 4. Grisi L, Leite RC, Martins JR, Barros AT, Andreotti R, Cancado PH, Leon AA, Pereira
521 JB, Villela HS. 2014. Reassessment of the potential economic impact of cattle
522 parasites in Brazil. Rev Bras Parasitol Vet 23:150-156.
523 5. Lane J, Jubb T, Shephard R, Webb-Ware J, Fordyce G. 2015. Priority list of endemic
524 diseases for the red meat industries. Meat & Livestock Australia.
525 6. Schnitzerling H, Noble P, Macqueen A, Dunham R. 1982. Resistance of the buffalo
526 fly, *Haematobia irritans exigua* (De Meijere), to two synthetic pyrethroids and DDT.
527 Aust J Entomol 21:77-80.
528 7. Rothwell J, Morgan J, James P, Brown G, Guerrero F, Jorgensen W. 2011.
529 Mechanism of resistance to synthetic pyrethroids in buffalo flies in south-east
530 Queensland. Aust Vet J 89:70-72.
531 8. Zug R, Hammerstein P. 2012. Still a host of hosts for *Wolbachia*: analysis of recent
532 data suggests that 40% of terrestrial arthropod species are infected. PLoS One
533 7:e38544.
534 9. LePage D, Bordenstein SR. 2013. *Wolbachia*: can we save lives with a great
535 pandemic? Trends Parasitol 29:385-393.
536 10. Caragata EP, Dutra HL, Moreira LA. 2016. Exploiting intimate relationships:
537 controlling mosquito-transmitted disease with *Wolbachia*. Trends Parasitol 32:207-
538 218.
539 11. Beckmann JF, Bonneau M, Chen H, Hochstrasser M, Poinsot D, Merçot H, Weill M,
540 Sicard M, Charlat S. 2019. The toxin-antidote model of cytoplasmic incompatibility:
541 Genetics and evolutionary implications. Trends Genet 35:175-185.
542 12. Werren JH, Baldo L, Clark ME. 2008. *Wolbachia*: master manipulators of invertebrate
543 biology. Nat Rev Micriobiol 6:741.
544 13. Poinsot D, Charlat S, Mercot H. 2003. On the mechanism of *Wolbachia*-induced
545 cytoplasmic incompatibility: Confronting the models with the facts. Bioessays 25:259-
546 265.
547 14. Landmann F, Orsi GA, Loppin B, Sullivan W. 2009. *Wolbachia*-mediated cytoplasmic
548 incompatibility is associated with impaired histone deposition in the male pronucleus.
549 PLoS Pathogens 5:e1000343.
550 15. Tram U, Sullivan W. 2002. Role of delayed nuclear envelope breakdown and mitosis
551 in *Wolbachia*-induced cytoplasmic incompatibility. Science 296:1124-1126.
552 16. Tram U, Fredrick K, Werren JH, Sullivan W. 2006. Paternal chromosome segregation
553 during the first mitotic division determines *Wolbachia*-induced cytoplasmic
554 incompatibility phenotype. J Cell Sci 119:3655-3663.
555 17. Callaini G, Dallai R, Riparbelli MG. 1997. *Wolbachia*-induced delay of paternal
556 chromatin condensation does not prevent maternal chromosomes from entering
557 anaphase in incompatible crosses of *Drosophila simulans*. J Cell Sci 110:271-280.
558 18. LePage DP, Metcalf JA, Bordenstein SR, On J, Perlmutter JI, Shropshire JD, Layton
559 EM, Funkhouser-Jones LJ, Beckmann JF, Bordenstein SR. 2017. Prophage WO
560 genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility.
561 Nature 543:243.

Accepted Manuscript Posted Online

562 19. Beckmann JF, Ronau JA, Hochstrasser M. 2017. A *Wolbachia* deubiquitylating
563 enzyme induces cytoplasmic incompatibility. Nat Microbiol 2:17007.
564 20. Shropshire JD, On J, Layton EM, Zhou H, Bordenstein SR. 2018. One prophage WO
565 gene rescues cytoplasmic incompatibility in *Drosophila melanogaster*. Proc Natl Acad
566 Sci, USA 115:4987-4991.
567 21. Jeffries CL, Walker T. 2016. *Wolbachia* biocontrol strategies for arboviral diseases
568 and the potential influence of resident *Wolbachia* strains in mosquitoes. Curr Trop
569 Med Rep 3:20-25.
570 22. Iturbe-Ormaetxe I, Walker T, O'Neill SL. 2011. *Wolbachia* and the biological control
571 of mosquito-borne disease. EMBO Rep 12:508-518.
572 23. Berticat C, Rousset F, Raymond M, Berthomieu A, Weill M. 2002. High *Wolbachia*
573 density in insecticide-resistant mosquitoes. Proc Biol Sci 269:1413-6.
574 24. Duron O, Labbe P, Berticat C, Rousset F, Guillot S, Raymond M, Weill M. 2006. High
575 *Wolbachia* density correlates with cost of infection for insecticide resistant *Culex*
576 *pipiens* mosquitoes. Evolution 60:303-14.
577 25. Echaubard P, Duron O, Agnew P, Sidobre C, Noel V, Weill M, Michalakis Y. 2010.
578 Rapid evolution of *Wolbachia* density in insecticide resistant *Culex pipiens*. Heredity
579 104:15-19.
580 26. Zhang B, McGraw E, Floate KD, James P, Jorgensen W, Rothwell J. 2009.
581 *Wolbachia* infection in Australasian and North American populations of *Haematobia*
582 *irritans* (Diptera: Muscidae). Vet Parasitol 162:350-353.
583 27. Torres L, Almazán C, Ayllón N, Galindo RC, Rosario-Cruz R, Quiroz-Romero H,
584 Gortazar C, de la Fuente J. 2012. Identification of microorganisms in partially fed
585 female horn flies, *Haematobia irritans*. Parasitol Res 111:1391-1395.
586 28. Palavesam A, Guerrero FD, Heekin AM, Wang J, Dowd SE, Sun Y, Foil LD, de León
587 AAP. 2012. Pyrosequencing-based analysis of the microbiome associated with the
588 horn fly, *Haematobia irritans*. PLoS One 7:e44390.
589 29. Jeyaprakash A, Hoy MA. 2000. Long PCR improves *Wolbachia* DNA amplification:
590 wsp sequences found in 76% of sixty-three arthropod species. Insect Mol Biol 9:393-
591 405.
592 30. Floate KD, Kyei-Poku GK, Coghlin PC. 2006. Overview and relevance of Wolbachia
593 bacteria in biocontrol research. Biocontrol Sci Technol 16:767-788.
594 31. Hornok S, Földvári G, Elek V, Naranjo V, Farkas R, de la Fuente J. 2008. Molecular
595 identification of *Anaplasma marginale* and rickettsial endosymbionts in blood-sucking
596 flies (Diptera: Tabanidae, Muscidae) and hard ticks (Acari: Ixodidae). Vet Parasitol
597 154:354-359.
598 32. Sutton ER, Harris SR, Parkhill J, Sinkins SP. 2014. Comparative genome analysis of
599 *Wolbachia* strain *w*Au. BMC Genomics 15:928.
600 33. Sinha A, Li Z, Sun L, Carlow CK. 2019. Complete genome sequence of the *Wolbachia*
601 *w*AlbB endosymbiont of *Aedes albopictus*. Genome Biol Evo 11:706-720.
602 34. Konganti K, Guerrero FD, Schilkey F, Ngam P, Jacobi JL, Umale PE, de Leon AAP,
603 Threadgill DW. 2018. A whole genome assembly of the Horn fly, *Haematobia irritans*,
604 and prediction of genes with roles in metabolism and sex determination. G3: Genes,
605 Genomes, Genetics 8:1675-1686.
606 35. Klasson L, Westberg J, Sapountzis P, Näslund K, Lutnaes Y, Darby AC, Veneti Z,
607 Chen L, Braig HR, Garrett R. 2009. The mosaic genome structure of the *Wolbachia*
608 *w*Ri strain infecting *Drosophila simulans*. Proc Natl Acad Sci, USA 106:5725-5730.
609 36. Ellegaard KM, Klasson L, Näslund K, Bourtzis K, Andersson SG. 2013. Comparative
610 genomics of *Wolbachia* and the bacterial species concept. PLoS Genetics
611 9:e1003381.
612 37. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin
613 W, Esser C, Ahmadinejad N. 2004. Phylogenomics of the reproductive parasite

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

Applied and Environmental Microbiology AEM Applied and Environmental Microbiology

Accepted Manuscript Posted Online

614 *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic
615 elements. PLoS Biol 2:e69.
616 38. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler
617 transform. Bioinformatics 26:589-595.
618 39. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome
619 assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595.
620 40. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng
621 Q, Wortman J, Young SK. 2014. Pilon: an integrated tool for comprehensive microbial
622 variant detection and genome assembly improvement. PLoS One 9:e112963.
623 41. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L,
624 Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome
625 annotation pipeline. Nucleic Acids Res 44:6614-6624.
626 42. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G,
627 Kriventseva EV, Zdobnov EM. 2017. BUSCO applications from quality assessments
628 to gene prediction and phylogenomics. Mol Biol Evol 35:543-548.
629 43. Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing genome assembly and
630 annotation completeness. Methods Mol Biol 1962:227-245.
631 44. Wang Y, Coleman-Derr D, Chen G, Gu YQ. 2015. OrthoVenn: a web server for
632 genome wide comparison and annotation of orthologous clusters across multiple
633 species. Nucleic Acids Res 43:W78-W84.
634 45. Hertig M. 1936. The rickettsia, *Wolbachia pipientis* (gen. et sp. n.) and associated
635 inclusions of the mosquito, *Culex pipiens*. Parasitology 28:453-486.
636 46. Lo N, Paraskevopoulos C, Bourtzis K, O'neill S, Werren J, Bordenstein S, Bandi C.
637 2007. Taxonomic status of the intracellular bacterium *Wolbachia pipientis*. Int J Syst
638 Evol Microbiol 57:654-657.
639 47. Lindsey AR, Bordenstein SR, Newton IL, Rasgon JL. 2016. *Wolbachia pipientis*
640 should not be split into multiple species: A response to Ramírez-Puebla et
641 al.,"Species in Wolbachia? Proposal for the designation of 'Candidatus Wolbachia
642 bourtzisii','Candidatus Wolbachia onchocercicola','Candidatus Wolbachia
643 blaxteri','Candidatus Wolbachia brugii','Candidatus Wolbachia taylori','Candidatus
644 Wolbachia collembolicola'and 'Candidatus Wolbachia multihospitum'for the different
645 species within Wolbachia supergroups". Syst Appl Microbiol 39:220.
646 48. Ramírez-Puebla ST, Servín-Garcidueñas LE, Ormeño-Orrillo E, de León AV-P,
647 Rosenblueth M, Delaye L, Martínez J, Martínez-Romero E. 2015. Species in
648 *Wolbachia*? Proposal for the designation of 'Candidatus *Wolbachia*
649 bourtzisii','Candidatus *Wolbachia* onchocercicola','Candidatus *Wolbachia*
650 blaxteri','Candidatus *Wolbachia* brugii','Candidatus *Wolbachia* taylori','Candidatus
651 *Wolbachia* collembolicola' and 'Candidatus *Wolbachia* multihospitum'for the different
652 species within *Wolbachia* supergroups. Syst Appl Microbiol 38:390-399.
653 49. Baldo L, Dunning Hotopp JC, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR,
654 Hayashi C, Maiden MC, Tettelin H, Werren JH. 2006. Multilocus sequence typing
655 system for the endosymbiont *Wolbachia pipientis*. Appl Environ Microbiol 72:7098-
656 110.
657 50. Bleidorn C, Gerth M. 2017. A critical re-evaluation of multilocus sequence typing
658 (MLST) efforts in *Wolbachia*. FEMS Microbiol Ecol 94:fix163.
659 51. Turelli M, Cooper BS, Richardson KM, Ginsberg PS, Peckenpaugh B, Antelope CX,
660 Kim KJ, May MR, Abrieux A, Wilson DA. 2018. Rapid global spread of *w*Ri-like
661 *Wolbachia* across multiple *Drosophila*. Curr Biol 28:963-971.e8.
662 52. Pietri JE, DeBruhl H, Sullivan W. 2016. The rich somatic life of *Wolbachia*. Microbiol
663 Open 5:923-936.
664 53. Oliveira MT, Barau JG, Junqueira ACM, Feijão PC, da Rosa AC, Abreu CF, Azeredo-
665 Espin AML, Lessinger AC. 2008. Structure and evolution of the mitochondrial

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology

Accepted Manuscript Posted Online

666 genomes of Haematobia irritans and *Stomoxys calcitrans*: the Muscidae (Diptera:
667 Calyptratae) perspective. Mol Phylogenet Evol 48:850-857.
668 54. Ding S, Li X, Wang N, Cameron SL, Mao M, Wang Y, Xi Y, Yang D. 2015. The
669 phylogeny and evolutionary timescale of muscoidea (Diptera: Brachycera:
670 Calyptratae) inferred from mitochondrial genomes. PLoS One 10:e0134170.
671 55. Bandi C, Anderson TJ, Genchi C, Blaxter ML. 1998. Phylogeny of *Wolbachia* in filarial
672 nematodes. Proc Biol Sci 265:2407-13.
673 56. Werren JH, Zhang W, Guo LR. 1995. Evolution and phylogeny of *Wolbachia*:
674 reproductive parasites of arthropods. Proc Biol Sci 261:55-63.
675 57. Gerth M, Bleidorn C. 2016. Comparative genomics provides a timeframe for
676 *Wolbachia* evolution and exposes a recent biotin synthesis operon transfer. Nat
677 Microbiol 2:16241.
678 58. Kyei-Poku G, Giladi M, Coghlin P, Mokady O, Zchori-Fein E, Floate K. 2006.
679 *Wolbachia* in wasps parasitic on filth flies with emphasis on *Spalangia cameroni*.
680 Entomol Exp Appl 121:123-135.
681 59. Domingues LN, Guerrero FD, Cameron C, Farmer A, Bendele KG, Foil LD. 2018.
682 The assembled transcriptome of the adult horn fly, *Haematobia irritans*. Data Brief
683 19:1933-1940.
684 60. Guerrero F, Dowd S, Nene V, Foil L. 2008. Expressed cDNAS from embryonic and
685 larval stages of the horn fly (Diptera: Muscidae). J Med Entomol 45:686-692.
686 61. Torres L, Almazán C, Ayllón N, Galindo RC, Rosario-Cruz R, Quiroz-Romero H, de
687 la Fuente J. 2011. Functional genomics of the horn fly, *Haematobia irritans* (Linnaeus,
688 1758). BMC Genomics 12:105.
689 62. Ribeiro JM, Debat HJ, Boiani M, Ures X, Rocha S, Breijo M. 2019. An insight into the
690 sialome, mialome and virome of the horn fly, *Haematobia irritans*. BMC Genomics
691 20:616.
692 63. Ramakodi MP, Singh B, Wells JD, Guerrero F, Ray DA. 2015. A 454 sequencing
693 approach to dipteran mitochondrial genome research. Genomics 105:53-60.
694 64. Gillings MR. 2017. Lateral gene transfer, bacterial genome evolution, and the
695 Anthropocene. Annals NY Acad Sci 1389:20-36.
696 65. Kent BN, Salichos L, Gibbons JG, Rokas A, Newton IL, Clark ME, Bordenstein SR.
697 2011. Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*)
698 determined by targeted genome capture. Genome Biol Evo 3:209-218.
699 66. Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the
700 agents of open source evolution. Nat Rev Micriobiol 3:722.
701 67. Newton IL, Bordenstein SR. 2011. Correlations between bacterial ecology and mobile
702 DNA. Curr Microbiol 62:198-208.
703 68. Espinoza-Gómez F, López-Lemus AU, Rodriguez-Sanchez IP, Martinez-Fierro ML,
704 Newton-Sánchez OA, Chávez-Flores E, Delgado-Enciso I. 2011. Detection of
705 sequences from a potentially novel strain of cell fusing agent virus in Mexican
706 *Stegomyia* (*Aedes*) *aegypti* mosquitoes. Arch Virol 156:1263-1267.
707 69. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. 2019. MAFFT-DASH:
708 integrated protein sequence and structural alignment. Nucleic Acids Res 47:W5-
709 W10.
710 70. Metcalf JA, Jo M, Bordenstein SR, Jaenike J, Bordenstein SR. 2014. Recent genome
711 reduction of *Wolbachia* in *Drosophila recens* targets phage WO and narrows
712 candidates for reproductive parasitism. PeerJ 2:e529.
713 71. Klasson L, Walker T, Sebaihia M, Sanders MJ, Quail MA, Lord A, Sanders S, Earl J,
714 O'neill SL, Thomson N. 2008. Genome evolution of *Wolbachia* strain *w*Pip from the
715 *Culex pipiens* group. Mol Biol Evol 25:1877-1887.

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology

Accepted Manuscript Posted Online

716 72. Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, Bhattacharyya A,
717 Kapatral V, Kumar S, Posfai J. 2005. The *Wolbachia* genome of *Brugia malayi*:
718 endosymbiont evolution within a human pathogenic nematode. PLoS Biol 3:e121.
719 73. Newton IL, Clark ME, Kent BN, Bordenstein SR, Qu J, Richards S, Kelkar YD, Werren
720 JH. 2016. Comparative genomics of two closely related *Wolbachia* with different
721 reproductive effects on hosts. Genome Biol Evo 8:1526-1542.
722 74. Siguier P, Pérochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the
723 reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32-D36.
724 75. Hibler CP. 1966. Development of *Stephanofilaria stilesi* in the horn fly. J Parasitol
725 52:890-898.
726 76. Kent BN, Bordenstein SR. 2010. Phage WO of *Wolbachia*: lambda of the
727 endosymbiont world. Trends Microbiol 18:173-181.
728 77. Bordenstein SR, Bordenstein SR. 2016. Eukaryotic association module in phage WO
729 genomes from *Wolbachia*. Nat Commun 7:13155.
730 78. Tanaka K, Furukawa S, Nikoh N, Sasaki T, Fukatsu T. 2009. Complete WO phage
731 sequences reveal their dynamic evolutionary trajectories and putative functional
732 elements required for integration into the *Wolbachia* genome. Appl Environ Microbiol
733 75:5676-5686.
734 79. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER:
735 a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16-
736 W21.
737 80. Biliske JA, Batista PD, Grant CL, Harris HL. 2011. The bacteriophage WORiC is the
738 active phage element in wRi of *Drosophila simulans* and represents a conserved
739 class of WO phages. BMC Microbiol 11:251.
740 81. Iturbe-Ormaetxe I, Burke GR, Riegler M, O'Neill SL. 2005. Distribution, expression,
741 and motif variability of ankyrin domain genes in *Wolbachia pipientis*. J Bacteriol
742 187:5136-5145.
743 82. Voronin DA, Kiseleva EV. 2007. Functional role of proteins containing ankyrin
744 repeats. Tsitologiia 49:989-999.
745 83. Ishmael N, Dunning Hotopp JC, Ioannidis P, Biber S, Sakamoto J, Siozios S, Nene
746 V, Werren J, Bourtzis K, Bordenstein SR, Tettelin H. 2009. Extensive genomic
747 diversity of closely related *Wolbachia* strains. Microbiology 155:2211-2222.
748 84. Siozios S, Ioannidis P, Klasson L, Andersson SG, Braig HR, Bourtzis K. 2013. The
749 diversity and evolution of *Wolbachia* ankyrin repeat domain genes. PLoS One
750 8:e55390.
751 85. Lindsey AR, Rice DW, Bordenstein SR, Brooks AW, Bordenstein SR, Newton IL.
752 2018. Evolutionary genetics of cytoplasmic incompatibility genes cifA and cifB in
753 prophage WO of *Wolbachia*. Genome Biol Evo 10:434-451.
754 86. Wang GH, Xiao JH, Xiong TL, Li Z, Murphy RW, Huang DW. 2013. High-efficiency
755 thermal asymmetric interlaced PCR (hiTAIL-PCR) for determination of a highly
756 degenerated prophage WO genome in a *Wolbachia* strain infecting a fig wasp
757 species. Appl Environ Microbiol 79:7476-7481.
758 87. Cooper BS, Ginsberg PS, Turelli M, Matute DR. 2017. *Wolbachia* in the *Drosophila*
759 *yakuba* complex: Pervasive frequency variation and weak cytoplasmic
760 incompatibility, but no apparent effect on reproductive isolation. Genetics 205:333-
761 351.
762 88. Cooper BS, Vanderpool D, Conner WR, Matute DR, Turelli M. 2019. *Wolbachia*
763 acquisition by *Drosophila yakuba*-Clade hosts and transfer of incompatibility loci
764 between distantly related *Wolbachia*. Genetics 212:1399-1419.
765 89. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
766 sequence data. Bioinformatics 30:2114-2120.

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

767 90. Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. 2011. ISsaga is an
768 ensemble of web-based methods for high throughput identification and semi-
769 automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol
770 12:R30.
771 91. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version
772 7: improvements in performance and usability. Mol Biol Evol 30:772-780.
773 92. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer.
774 Bioinformatics 27:1009-10.
775 93. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
776 throughput. Nucleic Acids Res 32:1792-1797.
777 94. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with
778 BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969-1973.
779 95. Guerrero FD, Dowd SE, Sun Y, Saldivar L, Wiley GB, Macmil SL, Najar F, Roe BA,
780 Foil LD. 2009. Microarray analysis of female- and larval-specific gene expression in
781 the horn fly (Diptera: Muscidae). J Med Entomol 46:257-70.
782

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

783 **Figure legends**

784 **Figure 1: The *Wolbachia* endosymbiont of *Haematobia irritans irritans w*Irr is related**
785 **to *Wolbachia* endosymbionts from *Drosophila* hosts**. Maximum clade credibility (MCC)
786 tree resulting from BEAST analyses of 79 concatenated recombination free gene loci of
787 supergroup A and B *Wolbachia* strains previously identified by Bleidorn and Gerth (2018)
788 resulting in an alignment of 49,68 bp. Posterior probability values are indicated at the nodes.
789 *w*Irr indicated by an arrowhead and branch lengths represent the genetic distances.

790 **Figure 2: The *w*Irr strain has undergone genome rearrangements compared to other**
791 **Supergroup A *Wolbachia* strains.** Genomes were compared using the MAFFT (v7)
792 algorithm. Dot plots of LAST comparisons under (threshold score = 39 E=8.4e-11 A) *w*Irr
793 genome compared to *w*Mel (Genbank ID: NC_002978.6) B) *w*Irr compared to *w*Ri (Genbank
794 ID: NC_012416), and C) *w*Ri compared to *w*Mel. Similarities in the forward orientation (red)
795 and similarities suggesting inversions (blue).

796 **Figure 3: Expansion of IS elements in *w*Irr genome is associated with a *cifB***
797 **homologue with limited transcriptional activity.** A) Schematic diagram of genomic loci in
798 *w*Irr associated with the IS5_ssgr_IS427 IS family identified by ISsaga and BLASTn hits
799 against the *w*Ha genome (Genbank ID: NC_021089.1) and the *w*Bm genome (Genbank ID:
800 CP034333.1). B) Transcriptional activity of the putative CifB homologue E0495_03245 was
801 explored through pooling RNA-Seq reads originating from all tissues and developmental
802 stages of all *H. i. irritans* libraries that were mapped to the *w*Irr genome. The resultant BAM
803 files were visualized with Integrated Genomics Viewer (IGV v 2.5.2). Forward mapped reads
804 are shown in red, reverse orientation reads are shown in blue. Light blue and red regions
805 indicate a mapping quality number of 0 (MQ0) which indicates that the read maps to multiple
806 regions on the genome.

807 **Figure 4: Gene order comparisons between WO prophages.** Reciprocal BLASTn
808 analyses of (A) Comparisons between WOVitA (Genbank ID: KX522565) and WOIrr, and
809 (B) Comparisons between WOMelB (Genbank ID: NC_002978.6) and WOIrr. Genomic loci
810 in WO prophages were analysed using Easyfig and matching loci with max *E*-value (0.001).
811 Regions of nucleotide identity are indicated by grey shading from 100-65%. Annotations of
812 genes are coloured based on automated NCBI annotation and manual PFAM protein
813 database curation. The predicted eukaryotic association module (EAM) is shown on
814 WOVitA.

815 **Figure 5: The *Wolbachia* endosymbiont *w*Irr has horizontally acquired a second**
816 **cytoplasmic incompatibility loci.** Maximum clade credibility (MCC) tree resulting from
817 BEAST analyses of A) *cifA* and B) *cifB* homologues with Type numbers as designated by

Accepted Manuscript Posted Online

818 Lindsey et al. (2018). Posterior probability values are indicated at the nodes. *w*Irr CI genes
819 indicated by arrowheads and branch lengths (genetic distances).

820 **Figure 6: Both cytoplasmic incompatibility loci are transcriptionally active in the**
821 ***Wolbachia* strain of *w*Irr.** Pooled RNA-Seq reads originating from all tissues and
822 developmental stages of all *H. i. irritans* libraries were mapped to identified CI loci in *w*Irr
823 genomes. Resultant BAM files were visualized with Integrated Genomics Viewer (IGV v
824 2.5.2). Forward mapped reads are shown in red, reverse orientation reads are shown in
825 blue. Light blue and red reads indicate a mapping quality number of 0 (MAPQ=0) which
826 indicates that the read maps to multiple regions on the genome.

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Accepted Manuscript Posted Online

827 **Table 1:** Genome features of complete supergroup A *Wolbachia* strains

Strain designation	*w*Irr	*w*Ri	*w*Au	*w*Mel	*w*Ha
Supergroup	A	A	A	A	A
Host	*H. i. irritans*	*D. simulans*	*D. simulans*	*D. melanogaster*	*D. simulans*
Genome size (Mb)	1.35	1.44	1.26	1.26	1.29
G+C (%)	35.3	35.2	35.2	35.2	35.3
Coding genes (Protein)	1,249	1,254	1,099	1,100	1,126
rRNA	3	3	3	3	3
tRNA	34	34	34	34	34
Other RNA	4	4	4	4	4
Prophage regions	5	4	3	3	2
Total genes	1,419	1403	1265	1270	1263
Total pseudo genes (%)	129	108	125	129	95
BUSCO score (%)	82.4	81.9	81.9	81.9	81.4
Reference	This study	(35)	(32)	(37)	(36)

828

829

830

831 **Table 2:** Transcriptional activity of *w*Irr in all life stages of *H. i. irritans* Kerrville colony

Run Accession	Sample	Reads Total	Reads mapped to *w*Irr	Transcripts per million
SRR6231662	Malpighian tubules	34,421,850	60,443	1,755.94
SRR6231666	Salivary gland	27,461,942	2,580	93.94
SRR6231664	Adult 0h post feed	44,820,832	181,392	4,047.04
SRR6231671	Adult 24h post feed	47,070,858	188,823	4,011.46
SRR6231665	Adult 2h post feed	51,829,048	827,413	15,964.27
SRR6231670	Adult 4h post feed	50,512,062	205,766	4,073.60
SRR6231654	Egg 0h	34,417,392	34,021	988.48
SRR6231655	Egg 2h	38,158,656	48,477	1,270.40
SRR6231660	Egg 4h	31,907,062	64,611	2,024.97
SRR6231661	Egg 9h	33,832,868	61,261	1,810.69
SRR6231658	Midgut	33,327,082	26,281	788.57
SRR6231659	Legs	38,066,002	175,711	4,615.95
SRR6231663	Ovary	39,372,642	51,187	1,300.06
SRR6231668	Pupae 1d	53,975,828	274,698	5,089.27
SRR6231669	Pupae 3d	51,859,710	377,067	7,270.90
SRR6231667	Testes	87,918,073	1,167,195	13,275.93

832

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

Accepted Manuscript Posted Online

833 **Table 3:** Cytoplasmic incompatibility (CI) genes identified in *w*Irr and closely related protein

Gene	Size	Position	Top Blastp hit, Query cover, Percentage identity, GenbankID
E0495_02160 cidA	483aa	414,360- 415,811	cidA *w*Pip, 100%, 82.89% AGR50404.1
E0495_02165 cifB	1132aa	415,858- 419,216	cifB *w*Ha, 99%, 91.81%, WP_144054595.1
cifA E0495_02270	474aa	441,815- 443,239	cifA *w*Mel 100%, 99.79%, WP_044471237.1
cifB E0495_02275	1166aa	443,315 – 446,815	cifB *w*Mel, 99%, 99.40%, AYE93038.1
E0495_03245	546aa	629,487 – 631,127	cifB *w*Ha, 98% 65.71%WP_144054595.1

834

835

836 **Table 4:** Metadata of available sequencing data of *H. i. irritans* samples

Location of *H. i. irritans* sample & collection date	Sample & type of sequencing	NCBI accession	Reference
Agricultural Center St. Gabriel Research Station Louisiana, USA Collected: 2008	Larval and embryonic samples, PolyA enriched RNA-Seq EST	Larval EST: FD457983-FD466257 Embryonic EST: FD449556-FD457982	(60)
Agricultural Center St. Gabriel Research Station Louisiana, USA Collected: 28 July 2010	Whole male and females. Permethrin treated surviving males and Permethrin + Piperonyl Butoxide treated killed males PolyA enriched RNA-Seq Illumina Genome Analyzer II/ Illumina HiSeq 2000	Assembled transcriptome accession: GGLM01000000 Bioproject accession: PRJNA429442	(59)
Agricultural Center St. Gabriel Research Station Louisiana, USA Collected: 2010	Eggs, larvae, whole male and females. PolyA enriched RNA-Seq 454	Male: SRR003192 Female: SRR003191 Egg: SRR003190 Larvae: SRR003189	(95)
Ciudad Victoria, Tamaulipas, Mexico Collected: prior to August 2010	Abdominal tissues of partially fed adult female, PolyA enriched RNA-Seq EST	HO000420-HO001165 HO004499-HO004744	(61)
Pressler Cattle Ranch Kerrville, Texas, USA Originally collected: 2003 Sampled: 2010	Single male adult, Random DNA sequenced using 454	SRA: SRR1578740	(63)
Canelones, Uruguay Collected: 2016	Salivary glands and midgut samples, PolyA enriched RNA-Seq Illumina HiSeq 2000	SRA Salivary glands: SRR5136552, SRR5136553 SRA midguts: SRR5136554, SRR5136555	(62)

837

838

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

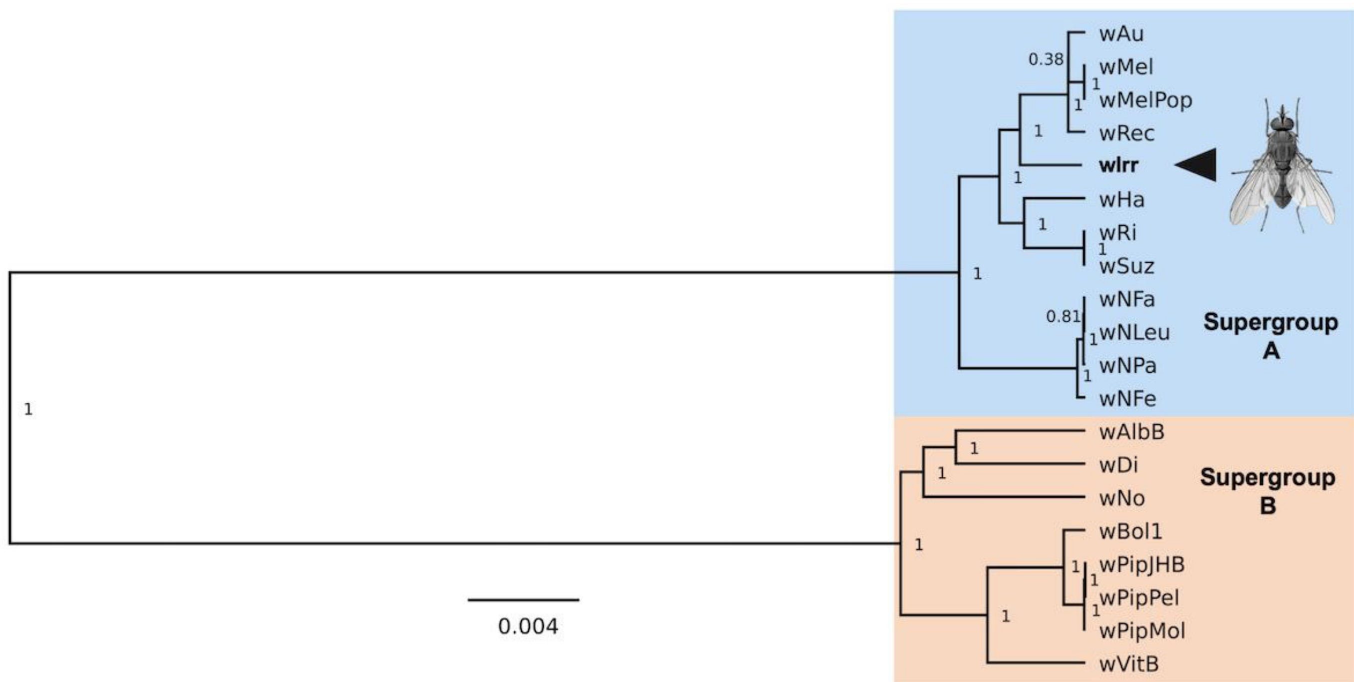
Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology



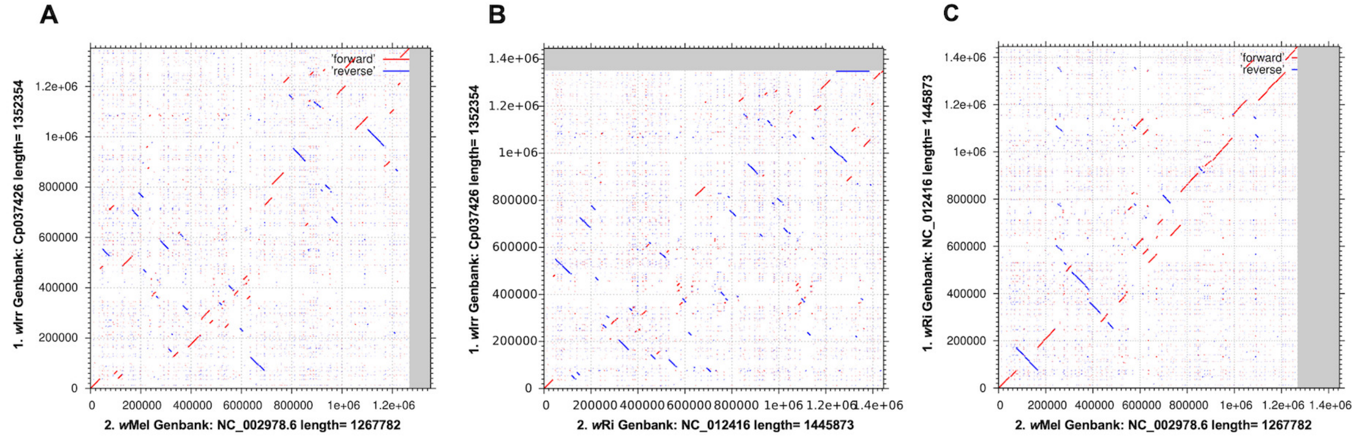
Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology



Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

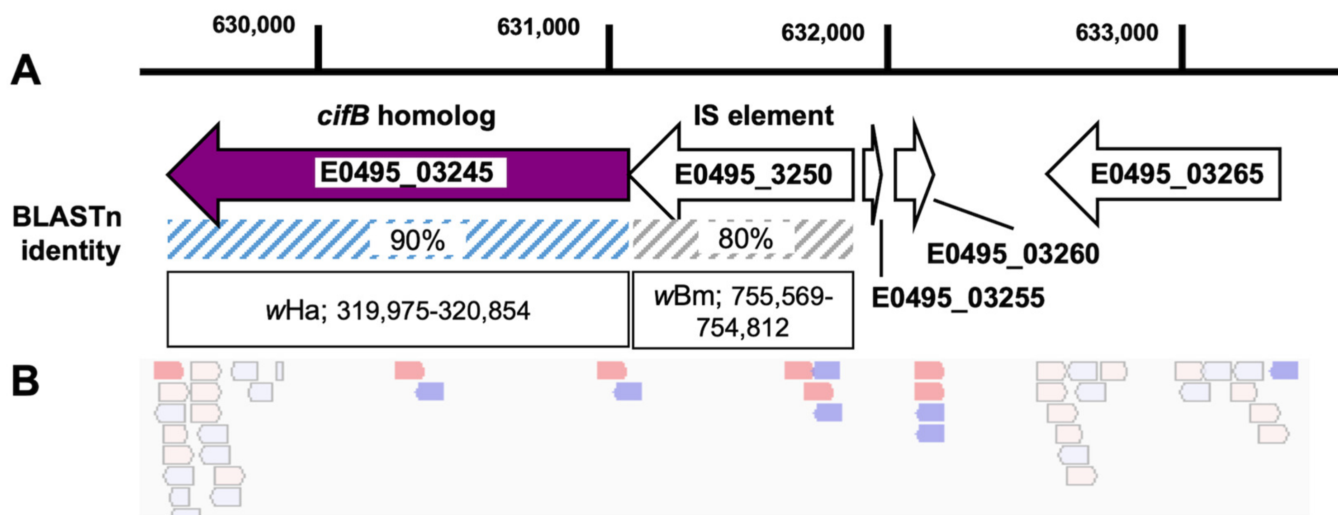
Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

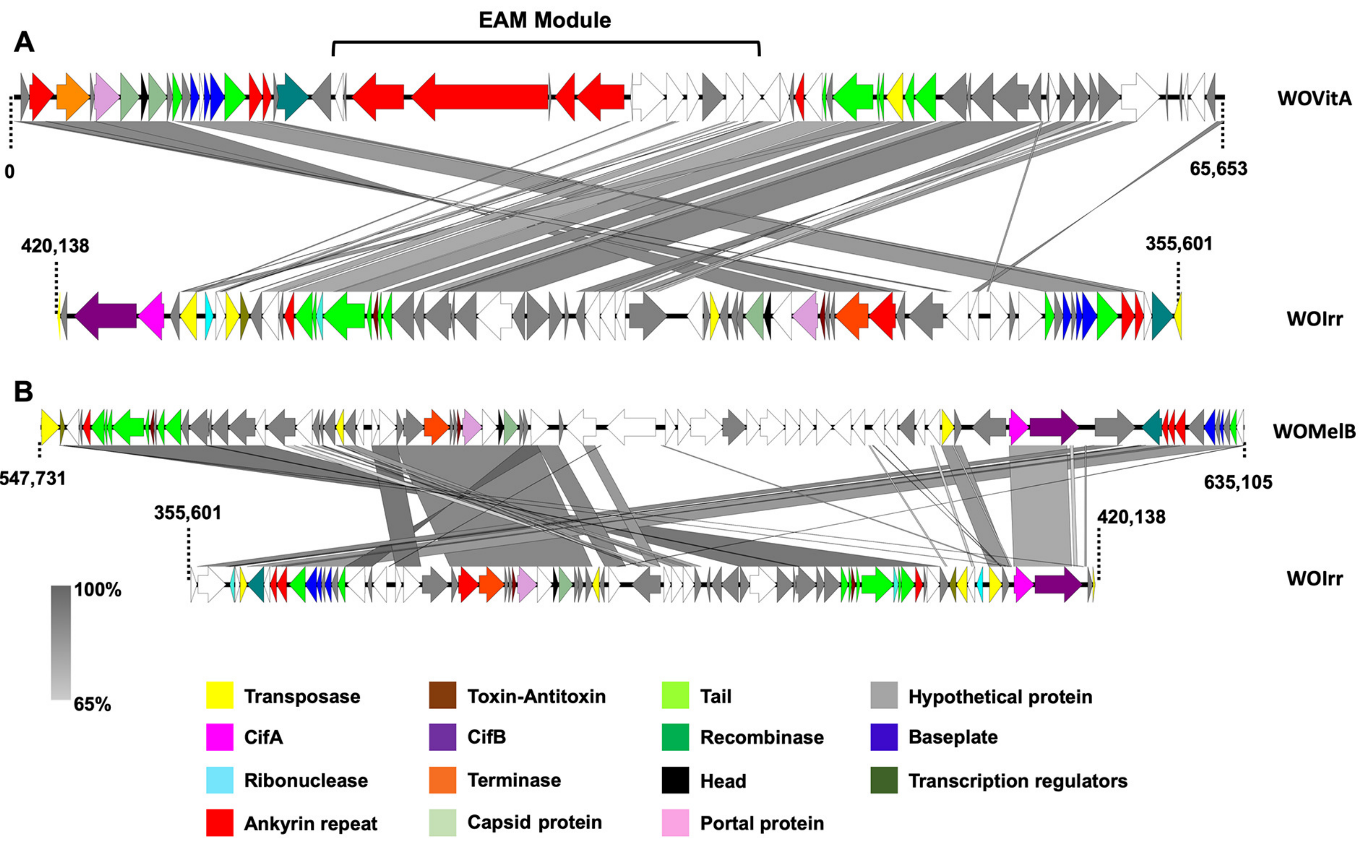


Accepted Manuscript Posted Online

Applied and Environmental Microbiology

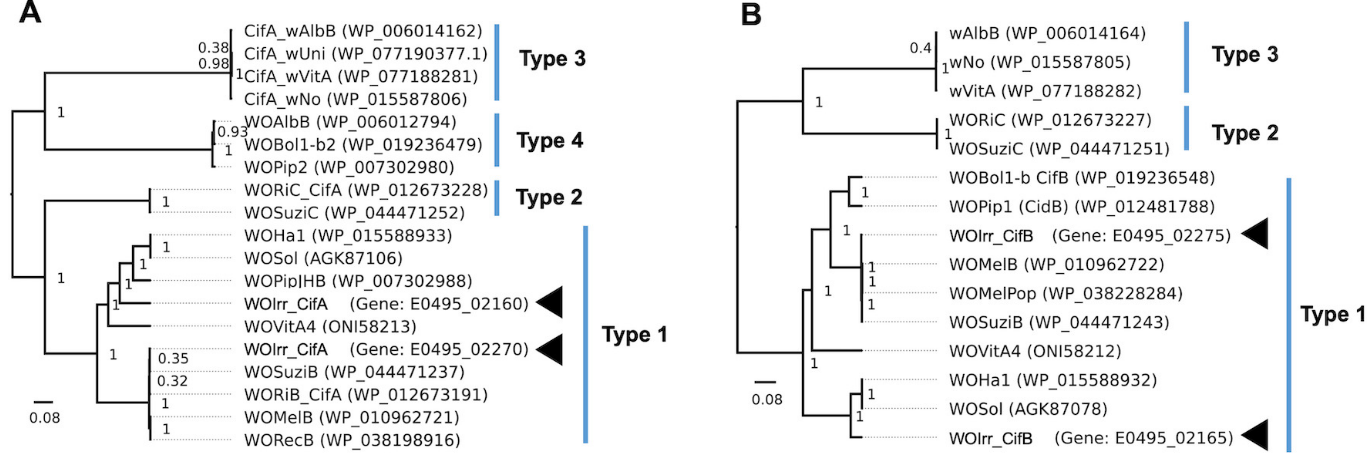
AEM

Applied and Environmental Microbiology



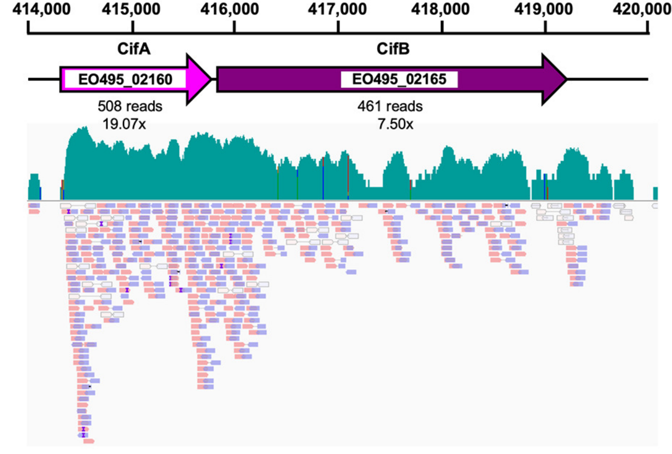
Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

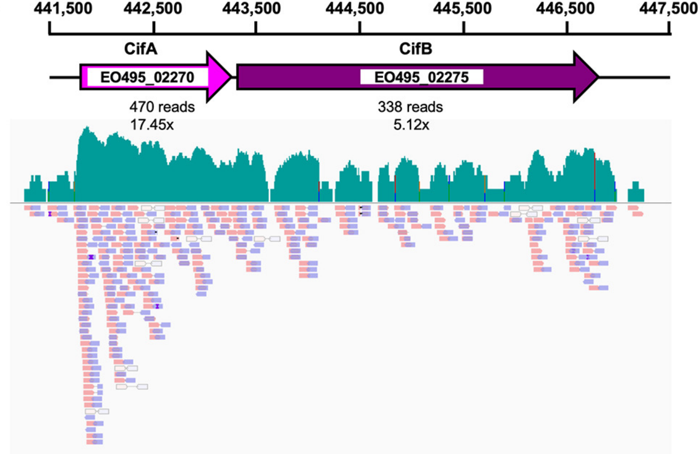
Accepted Manuscript Posted Online
Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology



Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library

Accepted Manuscript Posted Online
Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology





Downloaded from http://aem.asm.org/ on January 6, 2020 at University of Queensland Library