Male-killing-associated bacteriophage WO identified from comparisons of *Wolbachia* endosymbionts of *Homona magnanima*

Hiroshi Arai[1,2#*], Hisashi Anbutsu[3,4,5#], Yohei Nishikawa[3,5 #], Masato Kogawa[5], Kazuo Ishii[1], Masahito Hosokawa[3,5,6,7], Shiou-Ruei Lin[8], Masatoshi Ueda[1], Madoka Nakai[1], Yasuhisa Kunimi[1], Toshiyuki Harumoto[9,10], Daisuke Kageyama[2], Haruko Takeyama[3,5,6,7*], Maki N. Inoue[1*]

[1] United Graduate School of Agricultural Science, Tokyo University of Agriculture and Technology, 3-5-8 Saiwai, Fuchu, Tokyo 183-8509, Japan

[2] National Agriculture and Food Research Organization (NARO), 1-2 Owashi, Tsukuba, Ibaraki, 305-0851, Japan

[3] Computational Bio Big-Data Open Innovation Laboratory (CBBD-OIL), National Institute of Advanced Industrial Science and Technology (AIST), 3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555, Japan

[4] Bioproduction Research Institute, AIST, 1-1-1 Higashi, Tsukuba, Ibaraki, 305-8566, Japan

[5] Research Organization for Nano and Life Innovation, Waseda University, 513 Wasedatsurumaki-cho, Shinjuku-ku, Tokyo, 162-0041, Japan

[6] Institute for Advanced Research of Biosystem Dynamics, Waseda Research Institute for Science and Engineering, Graduate School of Advanced Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555, Japan

[7] Department of Life Science and Medical Bioscience, Graduate School of Advanced Science and Engineering, Waseda University, 2-2 Wakamatsucho, Shinjuku-ku, Tokyo, 162-8480, Japan

[8] Tea Research and Extension Station, 326011 Chung-Hsing RD., Yangmei, Taoyuan, Taiwan, R.O.C.

[9] Hakubi Center for Advanced Research, Kyoto University. Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501 JAPAN

[10] Graduate School of Biostudies, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku,

Kyoto 606-8501, Japan

# The authors contributed equally to this work
* Co-correspondence:
Hiroshi Arai
United Graduate School of Agricultural Science, Tokyo University of Agriculture and Technology, 3-5-8 Saiwai, Fuchu, Tokyo 183-8509, Japan
Email: dazai39papilio@gmail.com

Haruko Takeyama
Research Organization for Nano and Life Innovation, Waseda University, 513 Wasedatsurumaki-cho, Shinjuku-ku, Tokyo, 162-0041, Japan
Email: haruko-takeyama@waseda.jp

Maki Inoue
United Graduate School of Agricultural Science, Tokyo University of Agriculture and Technology, 3-5-8 Saiwai, Fuchu, Tokyo 183-8509, Japan
Email: makimaki@cc.tuat.ac.jp

Abstract

The origin and mechanism of male-killing, an advantageous strategy employed by maternally transmitted symbionts such as *Wolbachia*, remain unclear. We compared genomes of four *Wolbachia* strains derived from *Homona magnanima*, a male-killing strain *w*Hm-t (1.5 Mb), and three non-male-killing strains, *w*Hm-a (1.1 Mb), *w*Hm-b (1.3 Mb), and *w*Hm-c (1.4 Mb). A *w*Hm-t-specific 76-kbp prophage region harboured two tandemly arrayed WO-mediated killing (*wmk*) gene homologs (*wmk-1*/*wmk-2* and *wmk-3*/*wmk-4*). Of these, *wmk-1* or *wmk-3* killed almost all *Drosophila melanogaster* individuals when transgenically overexpressed. Dual expression of *wmk-3* and *wmk-4* killed all males and rescued females. We propose a novel hypothesis wherein horizontally transmitted proto-*Wolbachia* with a single *wmk* killed both sexes, and tandem duplication of *wmk* allowed an evolutionary transition to a vertically transmitted symbiont, causing male-killing. Our study highlights the bacteriophage as a critical driver of the evolution of male-killing and argues for a conserved male-killing mechanism in diverse insects.

67 **Introduction**

68 Insects harbour various endosymbiotic microbes, which occasionally interact with their
69 hosts in mutualistic or parasitic manner (1-3). Male-killing (MK), the death of male
70 offspring of infected female (4), is a reproductive manipulation caused by a diverse array
71 of intracellular bacteria (5), microsporidia (6), and certain RNA viruses (7-8). MK is
72 considered to have substantial ecological and evolutionary impact on insects as well as
73 microbes, and the molecular mechanisms underlying MK have attracted extensive
74 attention for decades (4, 9-12).

75 One of the best-studied MK endosymbionts is an alpha-proteobacterium
76 *Wolbachia*, infecting approximately 40–60% of insects (13) and nematodes (14). MK
77 *Wolbachia* strains have been identified from a variety of insect taxa which are not always
78 phylogenetically closely related (15-16). Moreover, even closely related *Wolbachia*
79 strains exhibit distinct phenotypes (17-20). *Wolbachia* is considered to have gained the
80 ability to manipulate host reproduction multiple times by acquiring the bacteriophage WO
81 (21-23). Indeed, the WO phage encodes cytoplasmic incompatibility (CI)-inducing
82 factors (CifA and CifB), which result in the inability of sperm and eggs to form viable
83 offspring (21-22, **Table 1**). Besides, the WO-mediated killing (*wmk*) gene, a putative
84 transcriptional regulator located in the WO phage, is considered a candidate gene for MK
85 because it induces a weak male lethality when transgenically overexpressed in *Drosophila*
86 *melanogaster* (23). However, the *wmk* gene was derived from a non-MK *Wolbachia*
87 (*w*Mel) naturally harboured by *D. melanogaster*, and the reason behind MK is not well
88 understood. Moreover, two *wmk* homologs (i.e., one derived from *w*Suz, non-MK
89 *Wolbachia* of *Drosophila suzukii*, and the other derived from *w*Rec, which does not induce
90 MK in the native host *Drosophila recens* but causes MK when transferred into *Drosophila*
91 *subquinaria* (24)) killed all males and females when transgenically overexpressed in *D.*
92 *melanogaster* (25). These results suggested multiple host-dependent functions of *wmk.*
93 Interestingly, female killing is mal-adaptive for maternally transmitted symbionts, and its
94 evolutionary significance is not well understood. The WO phages frequently exhibit
95 horizontal transfers among *Wolbachia* strains and are integrated into bacterial genomes
96 as a temperate phage (prophage). Therefore, the WO phage may have been a critical driver

97 of the evolution of *Wolbachia*-induced reproductive manipulations such as MK and CI.

98 How microbes have acquired the ability to cause MK (origin) and how they exert
99 the MK phenotype (mechanism) are key questions that remain unanswered (23-28). The
100 bacterium *Spiroplasma poulsonii* (Mollicutes) causes MK by expressing a toxin-encoding
101 *spaid* gene in *Drosophila melanogaster.* Nonetheless, *spaid* homologs have not been
102 identified from other known male-killers, including *Wolbachia* (26). Besides, the
103 expression of *Wolbachia*-induced MK depends on host species and *Wolbachia* genotypes
104 (24, 29-31). For instance, the transinfection of the MK strain *w*Inn in *Drosophila innubila*
105 does not cause MK in *D. melanogaster* and *D. simulans* (29). In contrast, a CI-inducing
106 *w*CauA strain in *Ephestia* (=*Cadra*) *cautella* exhibits MK in *Ephestia kuehniella* (30).
107 Since identical *Wolbachia* strains can cause different phenotypes depending on the host
108 genetic background, comparative analysis of *Wolbachia* strains showing distinct
109 phenotypes in a fixed host is an important approach to identifying MK gene(s).

110 Here, we focused on four *Wolbachia* strains, *w*Hm-t, *w*Hm-a, *w*Hm-b, and *w*Hm-
111 c, harboured by a tea pest tortrix moth *Homona magnanima* (Tortricidae, Lepidoptera)
112 (20, 32). *w*Hm-t (belonging to supergroup B) causes MK and is phylogenetically close to
113 *w*Hm-c, which does not cause MK. On the other hand, *w*Hm-a and *w*Hm-b (both
114 belonging to supergroup A) do not cause MK. Collectively, we explored the genetic
115 characteristics of male-killers and the evolutionary processes in the acquisition of the MK
116 phenotype.

117

118 **Results**

119 **Comparative genomics of *Wolbachia* strains in *H. magnanima***

120 The MK-inducing *w*Hm-t was purified from the matriline W^{T12} by ultracentrifugation.
121 Both PacBio RSII data (1.11 Gb, 110,635 reads, average read length: 7,603 bp) and
122 Illumina data (1.44 Gb, 4,798,918 reads, paired-end 150 bp) were used to generate a
123 polished 1,542,158 bp complete genome (**Table 2**). To construct genomes of *w*Hm-a,
124 *w*Hm-b, and *w*Hm-c, an isofemale line derived from Yakushima Island (normal sex ratio
125 for at least 30 generations, female ratio = 0.52, n = 1542) was subjected to single-cell
126 analysis. Random sequencing of 384 libraries using single cells purified from males of

127 the Yakushima matriline generated 6, 17, and 40 Illumina datasets for *w*Hm-a, *w*Hm-b,
128 and *w*Hm-c, respectively. Following the quality assessment, GridION generated data of
129 138 Mb for *w*Hm-a (44,000 reads, average read length: 3,144 bp), 158 Mb for *w*Hm-b
130 (48,000 reads, average read length: 3,281 bp), and 338 Mb for *w*Hm-c (128,000 reads,
131 average read length: 2,637 bp) by using six single-cell amplicons for each strain. Genome
132 assembling and polishing generated a complete genome of *w*Hm-c (1,447,668 bp) as well
133 as draft genomes of *w*Hm-a (1,174,342 bp, 20 contigs) and *w*Hm-b (1,302,323 bp, 9
134 contigs) (**Table 2**).

135 High level of genome rearrangements were observed between *w*Hm-t, *w*Hm-a,
136 and *w*Hm-b (< 50% homologies) (**Fig. 1A-B**). On the other hand, phylogenetically closely
137 related *w*Hm-t and *w*Hm-c showed a high level of syntenic conservation (> 92%
138 homology) in most parts of the genomes and were almost perfectly colinear with several
139 inversions (**Fig. 1C**). *w*Hm-t and *w*Hm-c shared identical genes (84.9%, n = 1210) and
140 homologs (> 90% identity, 94.9%, n = 1353) (**Extended Data Table 1**). Exclusively
141 present in *w*Hm-t but not in *w*Hm-a, *w*Hm-b, and *w*Hm-c were six protein clusters
142 constituting 14 proteins (**Extended Data Fig. 1**), which were, however, not possessed by
143 MK *Wolbachia* found in other insects such as *Hypolimnas bolina* and *Drosophila* (i.e.,
144 *w*Bol1b (18), *w*Bor (23), and *w*Rec (23, 24)). On the other hand, two protein clusters that
145 were absent in *w*Hm-t but present in *w*Hm-a, *w*Hm-b, and *w*Hm-c were hypothetical
146 proteins and mobile elements. None of the putative proteins encoded by the MK Partiti-
147 like virus in *H. magnanima* (7) or the MK protein Spaid encoded by the MK *Spiroplasma*
148 in *Drosophila* (25) were homologous to any region of the *w*Hm-t genome.

149

150 **Identification of a *w*Hm-t-specific prophage region WO*w*Hm-t76**

151 Notably, a 76-kb insertion (hereafter referred to as WO*w*Hm-t76), highly homologous to
152 the *Wolbachia* phage WO (**Extended Data Figs. 2, 3**), was present in the *w*Hm-t genome
153 (**Fig. 1D**) but not in *w*Hm-a, *w*Hm-b, and *w*Hm-c. Although several other prophage WO
154 regions were identified on the *w*Hm-t and *w*Hm-c genomes (**Extended Data Figs. 2, 3**),
155 further resequencing analysis confirmed that the WO*w*Hm-t76 region was conserved in
156 *w*Hm-t (n = 15; Taiwan) but not in *w*Hm-c (n = 9; Japan). Other than the WO*w*Hm-t76

157 region, deleted or inserted loci in the *w*Hm-c were IS5 family transposases (**Fig. 1E**). To
158 further confirm whether WO*w*Hm-t76 exists consistently in *w*Hm-t, the field-collected *H.*
159 *magnanima* that were either infected with *w*Hm-c or *w*Hm-t were subjected to PCR that
160 detected a sequence of the *cifB*-like gene (wHmt_13140) located in the WO*w*Hm-t76
161 region (**Fig. 1F**). All the Japanese *H. magnanima* derived from 18 populations
162 consistently infected with the non-MK *w*Hm-c were negative (116 females and 129
163 males) for the *cifB*-like gene. On the other hand, all the 32 MK matrilines established
164 from a single population of Taiwan were positive, and all the 13 non-MK matrilines
165 established from the same population were negative for the *cifB*-like gene. In addition,
166 *w*Hm-t typically showed higher titres than non-MK *w*Hm-c in *H. magnanima* (P = 0.003
167 by Steel–Dwass test; **Extended Data Fig. 4**), but *w*Hm-c did not cause MK even when
168 its density was incidentally comparable to or higher than that of *w*Hm-t.
169
170 **WO*w*Hm-t76 harboured and expressed virulence-associated genes**
171 The WO*w*Hm-t76 region was annotated to harbour 83 genes (wHmt_12340 to 13160),
172 which contained *Wolbachia* titre/virulence-associated gene homologs such as ankyrin
173 repeat-containing genes (n = 6, such as the *w*Hm-t-specific pk2 homolog wHmt_12420,
174 **Extended Data Fig. 1**), virulence associated Octomom genes (n = 7), the *cifB*-like gene
175 containing ankyrin repeats and papain-like peptidase domains (n = 1), and *wmk* homologs
176 containing helix-turn-helix domains (n = 4) (**Fig. 2A**). RNA-seq and PCR assays
177 confirmed that 51 genes in the WO*w*Hm-t76 region, including *cifB-like* and *wmk* genes,
178 were expressed during embryogenesis and in the ovary of *H. magnanima* (**Extended**
179 **Data Fig. 5**). Interestingly, the two tandemly arrayed *wmk* pairs in the WO*w*Hmt76 region
180 (**Fig. 2B-C**) were conserved (100% nucleotide identity for each gene) and expressed in
181 the MK *w*Bol1b (**Extended Data Fig. 5**). Besides, six *wmk* genes found in other *w*Hm-t
182 genomic loci were identical to those of non-MK *w*Hm-c (100% nucleotide identity for
183 each gene, **Fig. 2B**). Although *cifB* and *cifA* genes were tandemly arrayed in both CI-
184 inducing *w*Mel and *w*Pip strains (21), only *cifB*-like gene (wHmt_13140, **Fig. 2D**) was
185 found in the WO*w*Hm-t76 region. The Type IV *cifA* gene was conserved on their common
186 region of the *w*Hm-c (wHmc_11930) and *w*Hm-t (wHmt_13170) strains. The CI-inducing

187 *w*Hm-b encoded two tandemly arrayed Type I and Type II *cifA* and *cifB* gene pairs.
188 Unexpectedly, *w*Hm-a, which does not induce CI, also possessed a tandemly arrayed Type
189 I *cifA* and *cifB* gene pair.
190
191 **Expression of some *wmk* genes singly killed both males and females in *Drosophila***
192 To assess whether the four *wmk* genes and a *cifB*-like gene on the WO*w*Hm-t-76 region
193 affect insect viability, we used the *GAL4/UAS* system in *D. melanogaster*, as described
194 by Harumoto and Lemaitre (26) and Perlmutter et al. (23), to zygotically overexpress each
195 of the codon-optimised synthetic *wmk*-1 (561 bp), *wmk*-2 (361 bp), *wmk*-3 (897 bp), *wmk*-
196 4 (1006 bp), and *cifB-like* (3558 bp) genes (**Fig. 3A**). When *wmk*-1 was overexpressed,
197 only a small number of adult females (n = 52) and no males emerged in contrast to the
198 control (982 females and 1014 males). When *wmk*-3 was overexpressed, almost no adults
199 emerged (one female only) in comparison to the control (793 females and 888 males).
200 Overexpression of *wmk*-2, *wmk*-4, and *cifB*-like genes did not affect survival or sex ratio
201 compared to that in the controls (P = 0.89, 0.99, and 0.46 by Steel–Dwass test).
202
203 **Dual expression of *wmk* resulted in male-killing *in Drosophila***
204 To assess whether the dual expression of tandemly arrayed *wmk* pairs has additive or
205 different effects on host lethality, we simultaneously overexpressed each pair of the *wmk*
206 genes (i.e., *wmk*-1 and *wmk*-2 or *wmk*-3 and *wmk*-4) in *Drosophila*. Dual expression of
207 *wmk-1* and *wmk-2* did not affect the viability and generated a normal sex ratio (male:
208 47.4%, female: 52.6%). Moreover, dual expression of tandemly arrayed *wmk*-3 and *wmk*-
209 4 increased the viability of female flies but led to low male ratios (male: 12.5%, female:
210 87.5%, **Fig. 3, Extended Data Fig. 6**).
211
212 **Discussion**
213 In order to explore the origin and mechanism of MK in maternally transmitted symbionts,
214 in this study, we compared the genomes of four *Wolbachia* strains derived from *H.*
215 *magnanima* (a male-killing strain and three non-male-killing strains). We identified a 76-
216 kb prophage region WO*w*Hm-t76, which is consistently present in the MK *Wolbachia* of

217 *H. magnanima* but not in other *Wolbachia* strains that do not cause MK in a fixed host *H.*
218 *magnanima*. Strikingly, a tandemly arrayed *wmk* gene pair encoded by WO*w*Hm-t76
219 caused strong MK when they were transgenically co-expressed in *Drosophila* flies.

220 The expression of MK is hypothesised to be mediated by factors, such as tropism
221 or the expression level of genes present in MK *Wolbachia* (18). In *H. magnanima*, *w*Hm-
222 t causes MK when infecting the host at a high density (20). In the current study, *w*Hm-t
223 typically showed higher titres than non-MK *w*Hm-c in *H. magnanima*, but *w*Hm-c did not
224 cause MK even when its density was incidentally comparable to or higher than that of
225 *w*Hm-t. These results suggest that a high bacterial dose is not sufficient to induce MK in
226 *H. magnanima*. *Octomom* genes, which enhance *Wolbachia* virulence and proliferation in
227 *Drosophila* (33), on the *w*Hm-t-specific WO*w*Hm-t76 region may play a role in
228 upregulating *w*Hm-t titre in *H. magnanima* to secure stable expression of MK. The WO
229 phage possesses both a lytic and lysogenic nature and transfers genes laterally among
230 *Wolbachia* strains (34-39); in the current study, the WO*w*Hm-t76 sequences were found
231 in WO virions that were purified from the *w*Hm-t-infected host (**Extended Data Fig. 3**).
232 We speculate that after the divergence of Japanese and Taiwanese *H. magnanima*
233 populations, ancestral *w*Hm-c type *Wolbachia* in Taiwanese *H. magnanima* acquired the
234 WO*w*Hm-t76 region through integration of a certain WO phage, which gave rise to the
235 MK-inducing *w*Hm-t.

236 The *Wolbachia*-induced MK mechanism has remained unclear so far; however,
237 we found that tandemly arrayed *wmk* genes caused strong male lethality with recovered
238 female viabilities. The *Spiroplasma* endosymbiont in *D. melanogaster* recapitulates MK
239 via the Spaid toxin that damages the male's X chromosome using male-specific dosage
240 compensation machinery, which normalises the expression of genes on the X
241 chromosome between males (XY) and females (XX) (26). In contrast, *wmk*-1 and *wmk*-3
242 on the WO*w*Hm-t76 induced male as well as female lethality with different intensities in
243 *Drosophila*; this is in line with the finding by Arai et al. (20) that *w*Hm-t also killed
244 females when its titre was high in *H. magnanima*. The *wmk* gene (WD0626) derived from
245 non-MK *w*Mel is reported to induce a weak MK in *Drosophila* (approximately 30% of
246 males (23)). Moreover, *wmk* homologs of *w*Rec (MK in *D. subquinaria*) and *w*Suz (non-

247 MK in *D. suzukii*), which only have several substitutions with WD0626, caused complete
248 death (25). We speculate that the *wmk*-induced MK phenotype is represented by a
249 different tolerance level between the sexes against *wmk* genes (**Fig. 4A**). Like the CI
250 machinery of *Wolbachia*, the tandemly arrayed *wmk* genes might have combined effects,
251 wherein one *wmk* shows toxicity toward both sexes and the other *wmk* acts as its
252 suppressor. MK is considered to trigger diversification of the host sex determination
253 pathways (10-11, 40-41). These systems in insects are diverse; for instance, the female-
254 heterogametic system of Lepidoptera (including *H. magnanima*) and the male-
255 heterogametic system of Diptera (including *D. melanogaster*) do not share any sex-
256 determining genes known to date, except for the widely conserved *doublesex* gene. MK
257 *Wolbachia* targets the dosage compensation system to damage male X chromosomes in
258 *Drosophila* (42), whilst targeting the sex-determining gene cascade in *Ostrinia* and *H.*
259 *magnanima*, thereby inducing improper dosage compensation in males (43-46). *Wmk,*
260 assigned as a transcriptional regulator, may induce transcriptional effects leading to
261 defects in host development, such as embryonic lethality (46) in Lepidoptera (*H.*
262 *magnanima*) and Diptera (*Drosophila*) by a common but yet-unknown mechanism. The
263 multiplication of *wmk* genes in *Wolbachia* genomes may have contributed to the
264 diversification of sex-determining systems of arthropods.

265 Why does maternally transmitted *Wolbachia* possess female lethal gene, which
266 should bring about its own ruin? We propose a new hypothesis that a virulent *wmk* gene
267 (e.g., *wmk-1* or *wmk-3*) was passed on by a horizontally transmitted pathogenic bacterium
268 (i.e., ancestral *Wolbachia*) of insects. Rescue of females by the other *wmk* gene (e.g., *wmk-*
269 *2* or *wmk-4*) should have allowed female offspring of the infected individuals to survive.
270 This alteration *per se* can be regarded as the evolutionary transition of a horizontally
271 transmitted pathogen to a vertically transmitted endosymbiont, which led to the prominent
272 success of *Wolbachia,* as demonstrated by its current ubiquity. In general, virulent genes
273 often exert duplications and substitutions through strong selective pressures (47-48). For
274 the best-known reproductive phenotype CI, causative tandemly arrayed *cifB* and *cifA*
275 genes are conserved among CI-inducing *Wolbachia* (21, 39). Strikingly, the present study
276 demonstrated that non-CI-inducing *w*Hm-a strain also carried tandemly arrayed intact *cif*

genes in its genome. These data suggest that the presence of *cif* genes cannot be an indicator of CI expression, despite the commonly held view that both *cifB* and *cifA* are necessary and sufficient to induce CI (21-22, 39). Likewise, the host-killing function of *wmk* genes may be widely conserved among diverse *Wolbachia* strains, including MK and non-MK strains. Indeed, MK *w*Hm-t and MK *w*Bol1b shared identical tandemly arrayed *wmk* gene pairs, whereas other MK (described in Perlmutter et al. (23)) and non-MK *Wolbachia* (e.g., *w*Hm-a, *w*Hm-b, and *w*Hm-c) strains generally possess multiple *wmk* gene cassettes regardless of their phenotype. It is tempting to speculate about an intense evolutionary arms race between *Wolbachia* and its host, which might have occurred in the past. An excessive female-biased sex ratio acts as a strong selection pressure to evolve host suppressors against MK in various insects (40). If MK were suppressed quantitatively, the production of a higher amount of MK toxins achieved by duplicating *wmk* gene cassettes would cancel the MK suppression (**Fig. 4B**). Subsequently, the development of host suppressors (10-11, 41) would trigger an evolutionary arms race between *Wolbachia* and hosts, which would increase the number of *wmk* genes in the *Wolbachia* genome. WO phages (such as WO*w*Hm-t76), vectoring *wmk* genes, may have contributed to the evolutionary arms race.

In summary, our study highlights the bacteriophage as a critical driver of the evolution of MK *Wolbachia*. Considering that *Wolbachia* has moved between different insect species and WO phages have moved between different *Wolbachia* strains (34-39), we speculate that the *wmk* genes in WO*w*Hm-t76 may induce MK in Lepidoptera (*H. magnanima*) as well as Diptera (*Drosophila*). However, we should be aware of the possibility that some other genes in WO*w*Hm-t76 do cause MK in *H. magnanima*. Further functional validations of these genes will contribute to an integrated understanding of the MK mechanisms induced by *Wolbachia* in diverse insects. Our discoveries offer insights into the molecular machinery of reproductive manipulations as well as evolutionary associations of MK *Wolbachia*, WO phages, and their host insects.

Acknowledgments

We thank Professor Greg Hurst (Institute of Integrative Biology, University of Liverpool,

Liverpool, UK) and Dr. Katsuhiko Ito (Tokyo University of Agriculture and Technology, Tokyo, Japan) for their kind advice. We also thank Hachiman-jyu Tea Factory (Yakushima, Kagoshima) for samplings of *H. magnanima* with the help of Mr. Hiroki Ishii and Ms. Mayu Nishino.

We wish to acknowledge support from the Japan Society for the Promotion of Science (JSPS) Research Fellowships for Young Scientists [Grant Number 19J13123 and 21J00895 to H. Arai], The Hakubi Project of Kyoto University (to T. Harumoto), JST ERATO [Grant Number JPMJER1902 to T. Harumoto], Nagase Science and Technology Foundation (to T. Harumoto) and MEXT KAKENHI Grant-in-Aid for Scientific Research(S) [Grant Number 17H06158] for H. Takeyama.

Data Accessibility and Benefit-Sharing

Data Accessibility: The sequence read data were deposited in DDBJ under accession numbers PRJDB13119 (BioProject) and SAMD00445397 to SAMD00445400 (BioSample). *Wolbachia* genomes (contigs) are available on the DDBJ database under the following accession numbers: *w*Hm-a (BQXF01000001 - BQXF01000020), *w*Hm-b (BQXG01000001 - BQXG01000009), *w*Hm-c (AP025639), and *w*Hm-t (AP025638). Homologs of *wmk*, *cif*, and Octomom genes (*w*Hm-a, *w*Hm-b, *w*Hm-c, and *w*Hm-t) and synthetic construct of *wmk* and *cif* genes were deposited in DDBJ under accession numbers LC701647 to LC701694.

Benefits Generated: *H. magnanima* was collected from Tea Research and Extension Station (Taoyuan City, Taiwan) and imported with permission from the Ministry of Agriculture, Forestry and Fisheries (No. 27 - Yokohama Shokubou 891 and No. 297 - Yokohama Shokubou 1326). All collaborators are included as co-authors, and the results have been shared with the provider communities. More broadly, our group is committed to international scientific partnerships as well as institutional capacity building.

Author Contributions

H. Arai conducted field surveys, genome sequencing, and data analysis; designed fly experiments; and wrote the original manuscript. H. Anbutsu coordinated the single-cell

analysis and contributed to the entire discussion. Y. Nishikawa conducted the single-cell genome sequencing. M. Kogawa conducted the single-cell genome construction. K. Ishii helped with the genome construction. M. Hosokawa supported the entire single-cell analysis. SR. Lin organized the collection of insects in Taiwan and contributed to the discussion. M. Ueda collected Japanese *H. magnanima*. M. Nakai supported the entire experiments. Y. Kunimi sampled insects and contributed to the entire discussion. T. Harumoto designed and supported the fly experiments. D. Kageyama supported the fly experiments and revised the original manuscript. Lastly, H. Arai, H. Takeyama, and M.N. Inoue took responsibility for the decision to submit the manuscript for publication and managed the experiments and discussion.

Conflict of Interest

The authors declare that they have no conflicts of interest.

References

1. Sasaki, T. & Ishikawa, H. Production of essential amino acids from glutamate by mycetocyte symbionts of the pea aphid, *Acyrthosiphon pisum*. *J. Insect Physiol.* **41**, 41-46 (1995).
2. Anbutsu, H. et al. Small genome symbiont underlies cuticle hardness in beetles. *Proc. Natl Acad. Sci. U. S. A.* **114**, E8382-E8391 (2017).
3. Werren, J. H., Baldo, L. & Clark, M. E. *Wolbachia*: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* **6**, 741-751 (2008).
4. Hurst, L. D. The incidences and evolution of cytoplasmic male killers. *Proc. R. Soc. Lond. B* **244**, 91-99 (1991).
5. Hurst, G. D. D. et al. Invasion of one insect species, *Adalia bipunctata*, by two different male-killing bacteria. *Insect Mol. Biol.* **8**, 133-139 (1999).
6. Andreadis, T. G. & Hall, D. W. Significance of transovarial infections of *Amblyospora sp*. (microspora: Thelohaniidae) in relation to parasite maintenance in the mosquito Culex salinarius. *J. Invertebr. Pathol.* **34**, 152-157 (1979).
7. Fujita, R. et al. Late Male-Killing Viruses in Homona magnanima Identified as

Osugoroshi Viruses, Novel Members of Partitiviridae. *Front. Microbiol.* 11, 620623 (2020).

8. Kageyama, D., Yoshimura, K., Sugimoto, T. N., Katoh, T. K. & Watada, M. Maternally transmitted non-bacterial male killer in *Drosophila biauraria*. *Biol. Lett.* **13**, 20170476 (2017).
9. Majerus, M. E. N. & Hurst, G. D. D. Ladybirds as a model system for the study of male-killing symbionts. *Entomophaga* **42**, 13-20 (1997).
10. Charlat, S. et al. Extraordinary flux in sex ratio. *Science* **317**, 214-214 (2007).
11. Hayashi, M., Nomura, M. & Kageyama, D. Rapid comeback of males: evolution of male-killer suppression in a green lacewing population. *Proc. R. Soc. Lond. B* **285**(1877), 20180369 (2018).
12. Werren, J. H. & Beukeboom, L. W. Sex determination, sex ratios, and genetic conflict. *Annu. Rev. Ecol. Syst.* **29**, 233-261 (1998).
13. Zug, R. & Hammerstein, P. Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLOS ONE* **7**, e38544 (2012).
14. Ferri, E. et al. New insights into the evolution of *Wolbachia* infections in filarial nematodes inferred from a large range of screened species. *PLOS ONE* **6**, e20843 (2011).
15. Zhou, W., Rousset, F. & O'Neil, S. Phylogeny and PCR–based classification of *Wolbachia* strains using *wsp* gene sequences. *Proc. Biol. Sci.* **265**, 509-515 (1998).
16. Baldo, L. et al. Multilocus sequence typing system for the endosymbiont Wolbachia pipientis. *Appl. Environ. Microbiol.* **72**, 7098-7110 (2006).
17. Jiggins, F. M., Hurst, G. D. D., Dolman, C. E. & Majerus, M. E. N. High-prevalence male-killing *Wolbachia* in the butterfly *Acraea encedana*. *J. Evol. Biol.* **13**, 495-501 (2000).
18. Duplouy, A. et al. Draft genome sequence of the male-killing *Wolbachia* strain w Bol1 reveals recent horizontal gene transfers from diverse sources. *BMC Genomics* **14**, 1-13 (2013).
19. Metcalf, J. A., Jo, M., Bordenstein, S. R., Jaenike, J. & Bordenstein, S. R. Recent

397 genome reduction of *Wolbachia* in *Drosophila recens* targets phage WO and narrows
398 candidates for reproductive parasitism. *PeerJ* **2**, e529 (2014).
399 20. Arai, H., Lin, S. R., Nakai, M., Kunimi, Y. & Inoue, M. N. Closely Related Male-
400 Killing and Nonmale-Killing *Wolbachia* Strains in the Oriental Tea Tortrix *Homona*
401 *magnanima*. *Microb. Ecol.* **79**, 1011-1020 (2020).
402 21. LePage, D. P. et al. Prophage WO genes recapitulate and enhance *Wolbachia*-induced
403 cytoplasmic incompatibility. *Nature* **543**, 243-247 (2017).
404 22. Beckmann, J. F., Ronau, J. A. & Hochstrasser, M. A *Wolbachia* deubiquitylating
405 enzyme induces cytoplasmic incompatibility. *Nat. Microbiol.* **2**, 17007 (2017).
406 23. Perlmutter, J. I. et al. The phage gene *wmk* is a candidate for male killing by a bacterial
407 endosymbiont. *PLOS Pathog.* **15**, e1007936 (2019).
408 24. Jaenike, J. Spontaneous emergence of a new *Wolbachia* phenotype. *Evolution* **61**,
409 2244-2252 (2007).
410 25. Perlmutter, J. I., Meyers, J. E. & Bordenstein, S. R. A single synonymous nucleotide
411 change impacts the male-killing phenotype of prophage WO gene wmk. *eLife* **10**,
412 e67686 (2021).
413 26. Harumoto, T. & Lemaitre, B. Male-killing toxin in a bacterial symbiont of *Drosophila*.
414 *Nature* **557**, 252-255 (2018).
415 27. Harumoto, T., Anbutsu, H., Lemaitre, B. & Fukatsu, T. Male-killing symbiont
416 damages host's dosage-compensated sex chromosome to induce embryonic apoptosis.
417 *Nat. Commun.* **7**, 12781 (2016).
418 28. Reynolds, L. A., Hornett, E. A., Jiggins, C. D. & Hurst, G. D. D. Suppression of
419 *Wolbachia*-mediated male-killing in the butterfly *Hypolimnas bolina* involves a single
420 genomic region. *PeerJ* **7**, e7677 (2019).
421 29. Veneti, Z. et al. Loss of reproductive parasitism following transfer of male-killing
422 *Wolbachia* to *Drosophila melanogaster* and *Drosophila simulans*. *Heredity* **109**, 306-
423 312 (2012).
424 30. Sasaki, T., Kubo, T. & Ishikawa, H. Interspecific transfer of *Wolbachia* between two
425 lepidopteran insects expressing cytoplasmic incompatibility: a *Wolbachia* variant
426 naturally infecting *Cadra cautella* causes male killing in *Ephestia kuehniella*.

427 *Genetics* **162**, 1313-1319 (2002).
428 31. Hornett, E. A. et al. Evolution of male-killer suppression in a natural population.
429 *PLOS Biol.* **4**, e283 (2006).
430 32. Arai, H. et al. Multiple infection and reproductive manipulations of *Wolbachia* in
431 *Homona magnanima* (Lepidoptera: Tortricidae). *Microb. Ecol.* **77**, 257-266 (2019).
432 33. Chrostek, E. & Teixeira, L. Mutualism breakdown by amplification of *Wolbachia*
433 genes. *PLOS Biol.* **13**, e1002065 (2015).
434 34. Masui, S. et al. Bacteriophage WO and virus-like particles in *Wolbachia*, an
435 endosymbiont of arthropods. *Biochem. Biophys. Res. Commun.* **283**, 1099-1104
436 (2001).
437 35. Fujii, Y., Kubo, T., Ishikawa, H. & Sasaki, T. Isolation and characterization of the
438 bacteriophage WO from *Wolbachia*, an arthropod endosymbiont. *Biochem. Biophys.*
439 *Res. Commun.* **317**, 1183-1188 (2004).
440 36. Kent, B. N. & Bordenstein, S. R. Phage WO of *Wolbachia*: lambda of the
441 endosymbiont world. *Trends Microbiol.* **18**, 173-181 (2010).
442 37. Bordenstein, S. R. & Wernegreen, J. J. Bacteriophage flux in endosymbionts
443 (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Mol. Biol.*
444 *Evol.* **21**, 1981-1991 (2004).
445 38. Chafee, M. E., Funk, D. J., Harrison, R. G. & Bordenstein, S. R. Lateral phage transfer
446 in obligate intracellular bacteria (*Wolbachia*): verification from natural populations.
447 *Mol. Biol. Evol.* **27**, 501-505 (2010).
448 39. Lindsey, A. R. I. et al. Evolutionary genetics of cytoplasmic incompatibility genes
449 *cifA* and *cifB* in prophage WO of *Wolbachia*. *Genome Biol. Evol.* **10**, 434-451 (2018).
450 40. Jones, J. D. & Dangl, J. L. The plant immune system. *Nature* **444**, 323-329 (2006).
451 41. Hill, T., Unckless, R. L. & Perlmutter, J. I. Rapid evolution and horizontal gene
452 transfer in the genome of a male-killing *Wolbachia*. *bioRxiv* (2020).
453 42. Hornett, E. A., Kageyama, D. & Hurst, G. D. D. Sex determination systems as the
454 interface between male-killing bacteria and their hosts. *Proc. Biol. Sci.* **289**, 20212781
455 (2022).
456 43. Hamilton, W. D. Extraordinary sex ratios. A sex-ratio theory for sex linkage and

457 inbreeding has new implications in cytogenetics and entomology. *Science* **156**, 477-
458 488 (1967).
459 44. Harumoto, T., Fukatsu, T., & Lemaitre, B. Common and unique strategies of male
460 killing evolved in two distinct *Drosophila* symbionts. *Proc. Biol. Sci.* **285**, 20172167
461 (2018).
462 45. Sugimoto, T. N., & Ishikawa, Y. A male-killing *Wolbachia* carries a feminizing factor
463 and is associated with degradation of the sex-determining system of its host. *Biol. Lett.*
464 **8**, 412-415 (2012).
465 46. Sugimoto, T. N., Kayukawa, T., Shinoda, T., Ishikawa, Y., & Tsuchida, T. Misdirection
466 of dosage compensation underlies bidirectional sex-specific death in *Wolbachia*-
467 infected *Ostrinia scapulalis*. *Insect Biochem. Mol. Biol.* **66**, 72-76 (2015).
468 47. Fukui, T., Kawamoto, M., Shoji, K., Kiuchi, T., Sugano, S., Shimada, T., Suzuki, Y.,
469 & Katsuma, S. The endosymbiotic bacterium *Wolbachia* selectively kills male hosts
470 by targeting the masculinizing gene. *PLoS Pathog*. **11**, e1005048 (2015).
471 48. Arai, H. et al. Distinct Effects of the Male-Killing Bacteria Wolbachia and
472 Spiroplasma and a Partiti-Like Virus in the Tea Pest Moth, *Homona magnanima*.
473 *bioRxiv* (2022).
474

475 **Materials and Methods**

476 **Sampling of *H. magnanima***

477 To construct *Wolbachia* genomes, we used two *H. magnanima* lines: a MK line (W^{T12}),
478 infected with *w*Hm-t, and a normal line (Yakushima), triply infected with *w*Hm-a, *w*Hm-
479 b, and *w*Hm-c. For resequencing of *Wolbachia*, we used field-collected Japanese and
480 Taiwanese *H. magnanima* (**Extended Data Table 2**).
481

482 **Genome construction of *Wolbachia***

483 For sequencing of the *w*Hm-t genome, *Wolbachia* cells purified from the W^{T12} line were
484 used for DNA extraction following Duplouy et al. (18) and Iturbe-Ormaetxe et al. (49).
485 Extracted DNA was whole-genome-amplified (WGA) with the REPLI-g Mini Kit
486 (Qiagen) following the manufacturer's protocol and was sequenced using PacBio RSII

487 and Illumina (150 bp paired-end) platforms. PacBio RSII data were assembled using
488 Canu (50), and subsequent polishing using Illumina data with minimap2 (51) and Pilon
489 (52) generated the complete genome of *w*Hm-t. For genome sequencing of *w*Hm-a, *w*Hm-
490 b, and *w*Hm-c, *Wolbachia* cells purified from the Yakushima line were subjected to single-
491 cell analysis using MiSeq (75 bp paired-end) and GridION, as described by Chijiiwa et
492 al. (53) and Nishikawa et al. (54). The sequence data from GridION were assembled by
493 Canu (50) and were polished via Pilon (52) using MiSeq data. *Wolbachia* genomes were
494 annotated via DFAST (55). Phage WO infections were annotated with Phaster (56).
495 *Wolbachia wmk*, *cifA,* and *cifB* genes were annotated using BLAST searches and utilized
496 to construct phylogenetic trees using MEGA7 (57) by using maximum likelihood with
497 bootstrap re-sampling of 1,000 replicates. Homology between the *spaid* (26) versus
498 *w*Hm-t genes or genes of the Partiti-like viruses, OGV (7), versus *w*Hm-t genes were
499 analysed by BLAST searches.
500
501 **Resequencing analysis and detection of the WO*w*Hm-t76 region**
502 DNA extracted from *H. magnanima* harbouring *w*Hm-t (n = 15, Taiwan) and *w*Hm-c (n
503 = 9, Japan) was subjected to resequencing on an Illumina platform (150 bp paired-end).
504 Resequenced data, mapped to the *w*Hm-t genome (W^{T12}) using minimap2 (51), were
505 converted to consensus genomes using SAMtools (58), followed by synteny analysis
506 using GView (https://server.gview.ca). Homologies of the WO*w*Hm-t76 region and other
507 WO phages were visualized using Easyfig (59) and Mauve (60). The presence or absence
508 of the WO*w*Hm-t76 region in field-collected *H. magnanima* was assessed by PCR
509 amplifying a *cifB*-like gene (**Extended Data Table 3**) in the WO*w*Hm-t76 region with
510 the Emerald Amp Max Master mix (TaKaRa) at 94°C for 3 min, 35 cycles of 94°C for 30
511 sec, 62°C for 30 sec, 72°C for 3 min, and final extension at 72°C for 7 min.
512
513 **WO purification, observation, and phylogenetic analysis**
514 WO virions were purified as described by Fujita et al. (7). The recovered virions were
515 dialyzed using PBS and were loaded on membrane-coated transmission electron
516 microscopy grids (Nissin EM) and stained with 2% uranyl acetate solution (pH 7.0). The

517 samples were observed using a JEM1400 Plus (Joel). DNA extracted from purified
518 virions was subjected to PCR detection for genes of WO (*cifB*-like and minor capsid
519 protein [ORF7]), *Wolbachia* (*wsp*), and *Homona* (*β-actin*). Phylogenetic trees of WO
520 minor capsid protein genes of *Wolbachia* strains were estimated using MEGA7 (57) by
521 using maximum likelihood with bootstrap re-sampling of 1,000 replicates.
522
523 **Gene expression of *w*Hm-t**
524 Gene expression dynamics of *w*Hm-t in *H. magnanima* were assessed by RNA-seq and
525 RT-PCR using RNA extracted from W^{T12} adults or egg masses of 12, 36, 60, 84, and 108
526 hours post oviposition, as described by Arai et al. (61). Gene expression levels were
527 assessed using minimap2 (51) and iDep9 (http://bioinformatics.sdstate.edu/idep/).
528 Specific primers for *wmk* (*wmk-1*, *wmk-2*, *wmk-3*, and *wmk-4*) and *CifB* genes (**Extended**
529 **Data Table 3**) were used for RT-PCR. Expression levels of the *wmk* genes were further
530 tested using a female *Hypolimnas bolina* (collected from Ishigaki Is., Okinawa)
531 harbouring *w*Bol1b.
532
533 **Fly stocks and expression experiments**
534 The *wmk* and *cifB*-like genes in the WO*w*Hm-t76 region were synthesised with codon
535 optimisation for *D. melanogaster.* To express two tandemly arrayed *wmk* pairs, either (i)
536 *wmk*-1 and *wmk*-2 or (ii) *wmk*-3 and *wmk*-4 were conjugated using a T2A peptide as
537 described by Beckmann et al. (22). Transgenic fly lines were established using the
538 standard microinjection method for phiC31 integrase transformation into the third
539 chromosome at BestGene Inc. (Chino Hills, CA, UAS) using the synthetic constructs of
540 these genes in the pUASZ 1.1 plasmid. To transgenically overexpress the *wmk* and *cifB*-
541 like genes, *D. melanogaster* line *actin5C–GAL4* [*actin–GAL4*; BDSC 4414], cultured in
542 a tetracycline-containing medium (0.05% [w/v]) for one generation to eliminate native
543 *Wolbachia,* were crossed to homozygous *UAS* transgenic flies harbouring *wmk* or *cifB*-
544 like genes. The numbers of offspring expressing transgenic genes (*actin-GAL4/UAS*) and
545 non-expressing genes (*Cyo/UAS*) were all counted and were compared using Steel–Dwass
546 test in R software v4.0 (https://www.r-project.org/).

547

548 **Methods-only References**

549 49. Iturbe-Ormaetxe, I., Woolfit, M., Rancès, E., Duplouy, A. & O'Neill, S. L. A simple
550 protocol to obtain highly pure *Wolbachia* endosymbiont DNA for genome sequencing.
551 *J. Microbiol. Methods* **84**, 134-136 (2011).

552 50. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Phillippy, A.M. Canu: scalable and
553 accurate long-read assembly via adaptive k-mer weighting and repeat separation.
554 *Genome Res*. **27**, 722-736 (2017).

555 51. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**,
556 3094-3100 (2018).

557 52. Bruce, J. W. et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant
558 Detection and Genome Assembly Improvement. *PLoS ONE* **9**, e112963 (2014).

559 53. Chijiiwa, R. et al. Single-cell genomics of uncultured bacteria reveals dietary fiber
560 responders in the mouse gut microbiota. *Microbiome* **8**, 5 (2020).

561 54. Nishikawa, Y. et al. Massively parallel single-cell genome sequencing enables high-
562 resolution analysis of soil and marine microbiome. *bioRxiv* (2020).

563 55. Tanizawa, Y., Fujisawa, T., & Nakamura, Y. DFAST: a flexible prokaryotic genome
564 annotation pipeline for faster genome publication. *Bioinformatics* **34**, 1037-1039
565 (2018).

566 56. Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S.
567 PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids*
568 *Res.* **44**, W16-W21 (2016).

569 57. Kumar, S., Stecher, G., & Tamura, K. MEGA7: molecular evolutionary genetics
570 analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870-1874 (2016).

571 58. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**,
572 2078-2079 (2009).

573 59. Sullivan, M. J., Petty, N. K., & Beatson, S. A. Easyfig: a genome comparison
574 visualizer. *Bioinformatics* **27**, 1 (2011).

575 60. Darling, A. C., Mau, B., Blattner, F. R., & Perna, N. T. Mauve: multiple alignment of
576 conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394-1403

(2004).

61. Arai, H., Ishitsubo, Y., Nakai, M., & Inoue, M. N. Mass-Rearing and Molecular Studies in Tortricidae Pest Insects. *J. Vis. Exp.* **181** (2022).

Figure legends

Fig. 1 Genomic comparisons of *Wolbachia* strains in *H. magnanima*.

Contigs of *w*Hm-a (A), *w*Hm-b (B), and *w*Hm-c (C) were compared and sorted based on the complete genome of *w*Hm-t. (D) An insertion, the WO*w*Hm-t76 region, in the *w*Hm-t genome. The wHmc_11920 of the *w*Hm-c encodes the IS110 family of transposase and has identical sequences of wHmt_12330 of the *w*Hm-t. wHmc_11930 of *w*Hm-c, encoding cytoplasmic incompatibility factor A (CifA), is identical to wHmt_13170 of *w*Hm-t. (E) Genomic loci, absent from *w*Hm-c genomes (n = 9) but conserved in *w*Hm-t genomes (n = 15), are highlighted in red on the *w*Hm-t genome (1,542,158 bp). Homologous genes or loci between *w*Hm-c and *w*Hm-t, such as the WO*w*Hm-t76 region, are not highlighted in red. (F) Detection of the WO*w*Hm-t76 region from *Wolbachia* infecting Japanese and Taiwanese *H. magnanima*. F: female individuals, M; male individuals.

Fig. 2 Phenotype-associated genes in the WO*w*Hm-t76 region.

(A) Functions of genes in the WO*w*Hm-t76 region. (B) Phylogenetic trees of *wmk*. The *wmk* genes present in *w*Hm-t but absent in the *w*Hm-a, -b, and -c are highlighted with white diamonds (*wmk-1*, *wmk-2*, *wmk-3*, and *wmk-4*). Black diamonds indicate homologs of *wmk-1*, *wmk-2*, *wmk-3*, and *wmk-4* in the *w*Bol1b genome. (C) The domain structures of *wmk* annotated by InterPro. HTH: helix-turn-helix. (D) Phylogenetic trees of CifB homologs. Both *wmk* and *cifB* genes of *Wolbachia* strains were quoted from the NCBI database, and those of *w*Hm-a, -b, -c, and -t strains were manually annotated by local BLASTn and BLASTp in this study (Bit-score >100).

Fig. 3 Expression of *wmk* genes caused lethality in *Drosophila melanogaster*

The male ratio of adult progeny obtained from crosses between the *actin–GAL4* line and seven *UAS* transgenic lines (*wmk*-1, *wmk*-2, *wmk*-3, *wmk*-4, *cifB*-like, *wmk*-1 and *wmk*-2, and *wmk*-3 and *wmk*-4; n = 12-32 independent crosses for each transgene). We counted the number of resultant offspring (females, orange; males, blue) having both *actin–GAL4*

and *UAS* (+) and siblings having *UAS* and *Cyo (CyO)* as internal controls (−). The total number of adult counts for each genotype and sex are shown at the bottom. Different letters indicate statistically significant differences (Steel–Dwass test, $P < 0.05$). Dot plots show all data points individually. Exp.: expression; (-): non expressed; (+): expressed. Fem.: female progeny; Mal.: male progeny.

Fig. 4 Models for *wmk*-induced lethality and evolutionary arms race between insects and *Wolbachia*

(A) A model of *wmk* inducing lethality. The *wmk*-1 and *wmk*-3 cause MK and female-killing (FK) during development. Males and females show differences in tolerance to virulent *wmk* genes such as *wmk*-1 and *wmk*-3, leading to differences in MK or FK start points. High expression of virulent *wmk* genes leads to nearly complete death (all killing, AK). The co-expression of tandemly arrayed virulent *wmk* (such as *wmk*-1 and *wmk*-3) and non-killing (NK) *wmk* (such as *wmk*-2 and *wmk*-4) leads to low virulence towards their hosts. The thresholds of killing phenotypes exerted by *wmk* genes could also be different due to host genotype or species. (B) Evolutionary arms race between insects and *Wolbachia* triggered by *wmk*-induced pathogenicity. Our current view of the *wmk*-induced pathogenicity system can be represented by a 'zigzag' model. In phase 1, *Wolbachia* deploys the virulent *wmk* gene (*wmk-V*) to induce *wmk*-triggered virulence (WTV) that contributes to an all-killing phenotype (AK). In phase 2, natural selection (NS) drives *Wolbachia* to reduce virulence toward females by shedding the *wmk-V* gene through acquiring an additional *wmk* gene (*wmk-S*) that suppresses a given *wmk-V* function, which leads to the MK phenotype of *Wolbachia*. In phase 3, insects develop suppressor (Sup.) towards the MK phenotype either by direct disablement of *wmk* functions or develop the host's machinery to escape MK. In phase 4, duplications of *wmk* genes (*wmk-S2* and *wmk-V2*) result in MK, followed by the host's new resistance to MK. In phase 5, the WO phage transmits *wmk* gene cassettes from a *Wolbachia* to a new strain, which leads to MK in a new host insect.

Extended Data Fig. 1 Genetic comparisons of four *Wolbachia* strains in *H. magnanima*

Venn diagram of protein clusters of the *w*Hm-a, *w*Hm-b, *w*Hm-c, and *w*Hm-t strains. Numbers indicate gene clusters. Descriptions of *w*Hm-t-specific protein clusters (n = 6) and protein clusters specifically absent in the *w*Hm-t genome (n = 2) are highlighted in the table shown below the diagram.

Extended Data Fig. 2 Comparisons of WO in *w*Hm-t and *w*Hm-c.

(A) WO prophage regions were highlighted on both *w*Hm-t and *w*Hm-c genomes by PHASTER and manual BLASTn searches. (B) Homologies of WO regions on the *w*Hm-c and *w*Hm-t genomes. Blocks shown in the same colours indicate homologous genomic loci. The arrows indicate the WO region. Homologies are indicated with broken lines.

Extended Data Fig. 3 Phylogeny and morphology of WO.

(A) Phylogeny of *Wolbachia* phage WO, based on the minor capsid protein (ORF7). A maximum-likelihood tree (1,000 bootstraps) inferred from amino acid sites of *Wolbachia* strains is shown. Sequence accession numbers are indicated in the figure. I, II, and III represent the ORF7 clusters identified by Bordenstein and Wernegreen (37). (B) Purified WO phage particles visualized using a transmission electron microscope (TEM) with uranyl acetate staining. WO genes (ORF-7 and *cifB*-like) were detected from DNA extracted from purified WO particles (Pur) and *H. magnanima* harboring *w*Hm-t (W^{T12}). Both *Wolbachia* (*wsp*) and host (*b-actin*) were negative for the fraction. (C) Homologies of WO*w*Hmt-t76 and two *Wolbachia* phage WOs. The WOs identified from *w*Kue (AB036666) and *w*CauB (AB478515) were compared by BLASTn.

Extended Data Fig. 4 Titres of *Wolbachia* in Taiwanese *w*Hm-t-infected and Japanese *w*Hm-c-infected host individuals.

Wolbachia copies (*wsp*) in *w*Hm-t-infected females of Taiwanese host lines (collected by Arai et al., 2021) and *w*Hm-c-infected Japanese populations surveyed in this study were compared. The centreline within the box represents the median. The upper and lower

boundaries of the box indicate the upper quartile and lower quartile, respectively. The sample size (number of individuals) in each data point is displayed below the plots. Different letters indicate significant differences between groups (Steel-Dwass test, $p < 0.05$). MK: male-killing, NMK: non-male-killing, F: female. M: male.

Extended Data Fig. 5 Expression of genes in the WO*w*Hmt-t76 region.

(A) RT-PCR detection for the *wmk* and *cifB*-like genes. 12, 36, 60, 84, and 108 indicate hours post oviposition of *H. magnanima* egg masses. Ov+: female ovary of *H. magnanima*; Ov-: non-reverse-transcribed samples (RT minus, control) using RNA extracted from the female ovary of *H. magnanima*; Hb: female ovary of *Hypolimnas bolina*; P: DNA extracted from a W^{T12} adult female (positive control); N: negative control (water). (b) Expression of genes on the WO*w*Hmt-t76 region throughout embryogenesis. Mapping counts of each gene obtained by RNA-seq data (*H. magnanima* egg masses 12–84 hours post oviposition and 108-hour male or female embryos) were analysed. Expression levels are highlighted in blue (low) and red (high). F: female; M: male.

Extended Data Fig. 6 Number of female and male progenies in *D. melanogaster* expressing genes and their correlation

The number of adult progenies obtained from crosses between the *actin–GAL4* line and seven UAS transgenic lines: *wmk*-1 (A), *wmk*-2 (B), *wmk*-3 (C), *wmk*-4 (D), *cifB*-like (E), *wmk*-1 and *wmk*-2(F), and *wmk*-3 and *wmk*-4 (G) (n = 12-32 independent crosses for each transgene). We counted the number of resultant offspring (females, orange; males, blue) having both *actin–GAL4* and *UAS* (+) and siblings having *UAS* and *Cyo* as internal controls (−). The total number of adult counts for each genotype and sex are shown at the bottom. Exp.: expression; (-): non expressed; (+): expressed. (H) Dot plots of male and female viabilities in each independent cross for *wmk*-3 and *wmk*-4 dual expression.

Table 1 Reproductive phenotype-associated genes in *Wolbachia* and other male-killers

Species and strain	Host insects	*Wolbachia* phenotypes	Genes	Gene functions in transgenetically overexpressed *Drosophila*	References
Wolbachia pipientis *w*Pip	*Culex pipiens*	CI*[2]	*cifA* and *cifB*	Recapitulated CI phenotype.	Beckmann et al. (21)
Wolbachia pipientis *w*Mel	*Drosophila melanogaster*	CI*[2]	*cifA* and *cifB*	Recapitulated CI phenotype.	LePage et al. (22)
			wmk	Weak MK-like phenomenon (30% male death).	Perlmutter et al. (23)
Wolbachia pipientis *w*Rec	*Drosophila recens*	CI*[2]	*wmk*	All death	Jaenike (24) and Perlmutter et al. (25)
	*Drosophila subquinaria**[1]	MK*[3]			
Wolbachia pipientis *w*Suz	*Drosophila suzukii*	CI*[2]	*wmk*	All death	Perlmutter et al. (25)
Spiroplasma poulsonii (MSRO strain)	*Drosophila melanogaster*	MK*[3]	*spaid*	Recapitulated MK phenotype.	Harumoto and Lemaitre (26)

*1 *w*Rec induces MK in the offspring of *D. recens* and *D. subquinaria*. Susceptibility to MK is recessive.

*2 CI phenotype causes lethality in the offspring of *Wolbachia*-infected males and non-infected females.

*3 MK phenotype causes lethality in males and leads to a female-biased sex ratio.

Table 2 General characteristics of *Wolbachia* genomes

Strain	*w*Hm-a	*w*Hm-b	*w*Hm-c	*w*Hm-t
Supergroup	A	A	B	B
Phenotype	Not detected*[1]	CI*[1]	Shorten development time Increase pupal weight*[1]	Male-killing*[2]
Contigs	20	9	1	1
Genome length (nt)	1,174,342	1,302,323	1,447,668	1,542,158
GC Content (%)	35.28	35.12	34.01	34.06
CDS	1121	1,201	1,302	1,426
Repeat region	Not identified	200	128	177
tRNA	29	34	33	34
rRNA	3	3	3	3
Completeness (%)	85.3	98.0	99.7	100
Contamination (%)	3.1	2.7	0.3	0.3

*[1] Based on Arai et al. (2019).

*[2] Based on Arai et al. (2020).

Extended Data Table 1. Homologies of total 1,426 *w*Hm-t genes versus genes of other non-MK strains in *H. magnanima*

Homologies *[1]	*w*Hm-a	*w*Hm-b	*w*Hm-c
Genes with 100% identity	11	0	1,210
Genes with ≥ 95% identity	64	39	1,328
Genes with ≥ 90% identity	253	214	1,353
Genes with ≥ 80% identity	835	794	1,377
Genes with ≥ 70% identity	858	820	1,379
Genes with no homologies*[2]	568 (387)	606 (298)	47 (12)

*[1] The identity of genes was analysed by using BLASTn.

*[2] Numbers to the left of the parentheses indicate genes specific for *w*Hm-t, and numbers shown in parentheses indicate the genes specific for the compared non-MK strain.

Extended Data Table 2. Sampling information of Taiwanese and Japanese *H. magnanima*

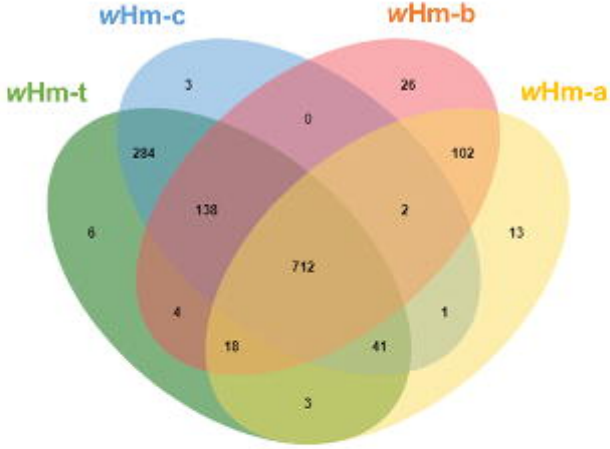
Country	Location	Male	Female	Month and Year	Remarks
Taiwan	Taoyuan, TRES	0	45	Oct. 2017	Arai et al., 2020, sequenced
Japan	Murakami, Niigata	12	12	Nov. 2019	This study
	Oku-kuji, Ibaraki	12	12	Nov. 2019	This study, sequenced
	Asahi, Toyama	12	12	May. 2019	This study, sequenced
	Fuchu, Tokyo	12	0	Jul. 2020	This study, sequenced
	Setagaya, Tokyo	7	0	May. 2020	This study, sequenced
	Mt. Takao, Tokyo	12	0	May. 2020	This study, sequenced
	Nirasaki, Yamanashi	2	0	Jul. 2020	This study, sequenced
	Shimada, Shizuoka	23	41	Sep. 2019	Takamatsu et al., 2021, sequenced
	Shimane, Shimane	3	0	Sep. 2021	This study
	Tsushima-Is., Nagasaki	0	1	Jun. 2018	This study
	Mimata, Miyazaki	0	8	Feb. 2010	This study
	Makurazaki, Kagoshima	0	8	Feb. 2010	This study
	Mizobe, Kagoshima	0	8	Feb. 2010	This study
	Chiran, Kagoshima	0	2	Feb. 2010	This study

Yakushima Is., Kagoshima	12	12	Mar. 2018	This study, sequenced
Tanegashima-Is., Kagoshima	4	0	Aug. 2019	This study, sequenced
Amami-Oshima Is., Kagoshima	13	0	Jun. 2020	This study, sequenced
Okinawa-honto Is., Okinawa	5	0	Nov. 2000	Morimoto et al., 2000

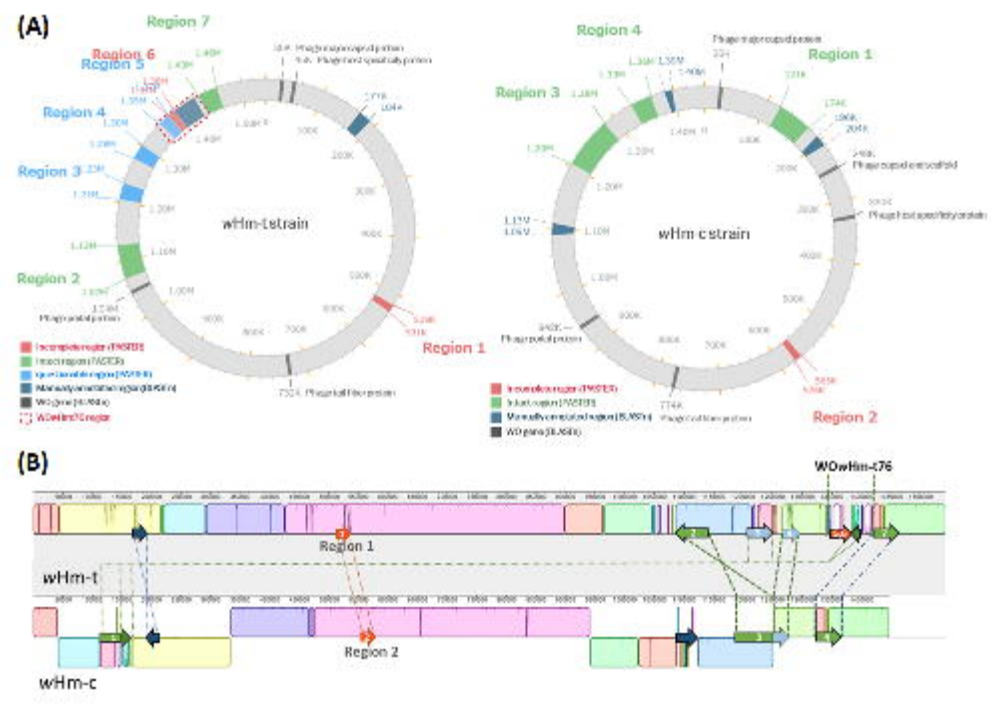
Extended Data Table 3. Primers for PCR detection.

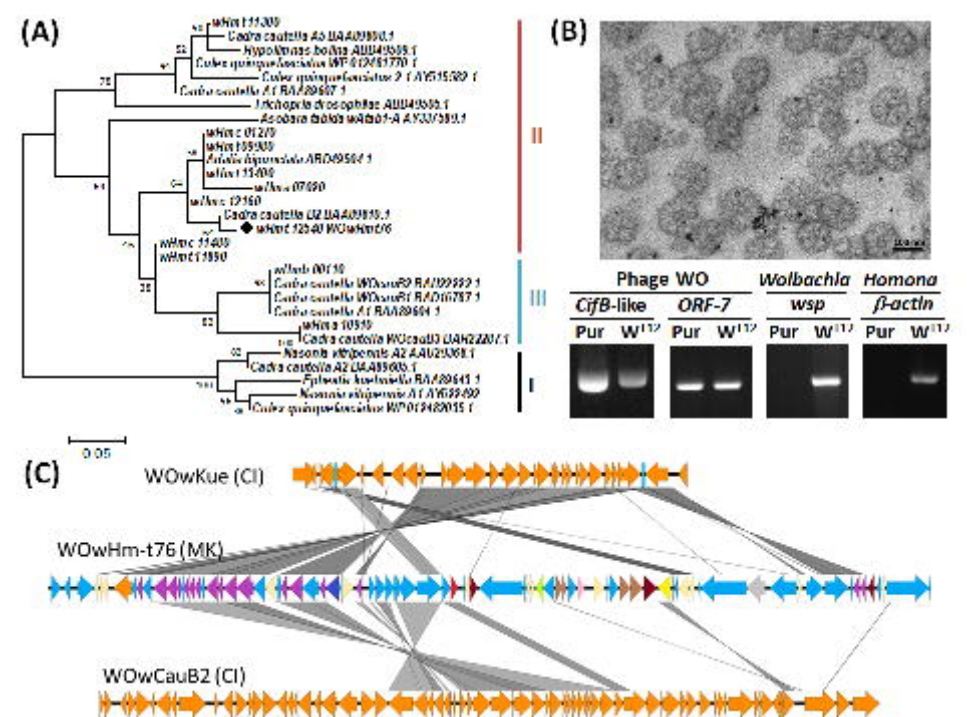
Target organisms	Gene	Primer sequences (5'-3')	Product size, bp	Annealing temperature, °C
Wolbachia	*Wsp*	wspF81：TGGTCCAATAAGTGATGAAGAAAC wspR691：TGGAGTAGCGTTTAATTTTT	611	60
Wolbachia (*w*Hm-a)	*Wsp*	wHm-a F321：CCTAAACAAAAATAATGTTACAG wHm-a R565：TTTGATCATTCACAGCGT	244	50
Wolbachia (*w*Hm-b)	*Wsp*	wHm-b F176：GGTGCTAAAAAGAAGACTGCGG wHm-b R667：CCCCCTTGTCTTTGCTTGC	491	60
Wolbachia (*w*Hm-c, -t)	*Wsp*	wHm-c F188：CATATAAATCAGGTAAGGACAAC wHm-c R603 ：CACCAGCTTTTGCTTGATA	415	55
Wolbachia (*w*Hm-a, b, -c, and -t)	*Wsp*	wHm-uni_qpcrF: TGGTGTTGGTGCAGCGTAT wHm-uni_qpcrR: AACTAACACCAGCTTTTGCTTGA	120	60
Wolbachia (*w*Hm-t)	WO*w*Hm-t76_CifB	WHMT_00358_f: ATGATTGAAGATAGAAATGTTCCTTTATCC WHMT_00358_r: CTACCTACCGCCTTTACCTTTGCTA	3545	62

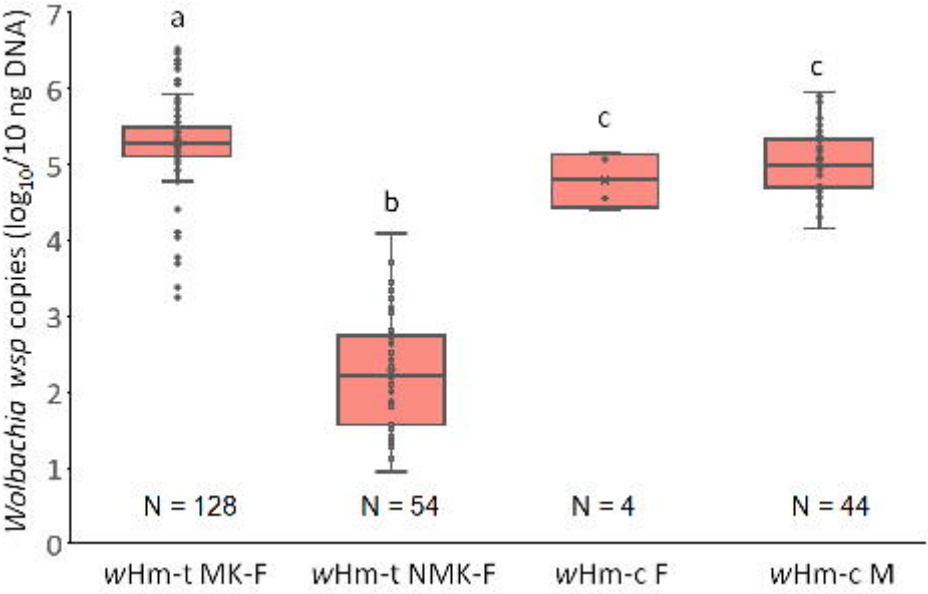
	wmk-1	wmk1F: ATGGCAAATATCTCTATAAGGT wmk1R: TCACGTGTTGTACTCATTTTTT	561	55
	wmk-2	wmk2F: ATGAGTACAACACGTGATGAAAT wmk2R: TAAATTTGTTGAATTTCGTCTAAAG	360	55
	wmk-3	wmk3F: GTGGAAAGAAGTTTAGACTATGAAGTAGGG wmk3R: TCCAATTTGATCAGCAGATAAATCAAC	1005	55
	wmk-4	wmk4F: ATGTTTGTTTCTGTAAGAGATGTTAGT wmk4R: TCAATTTTCACACTCACCCAACTCACCA	1089	55
WO phage	Minor capsid protein	orf7f: GAAATGCTTGTTCAGCTAATAGC orf7r: ATAAATTCTCCTATTTTTTCTGG	358	50
H. magnanima	*b-actin*	297f：AACTGGGATGACATGGAGAAGATCTGGC 1139r: GAGATCCACATCTGCTGGAAGGTGGACAG	838	55
	Lepidoptera COI	LepF: ATTCAACCAATCATAAAGATATTGG LepR: TAAACTTCTGGATGTCCAAAAAATCA	650	55

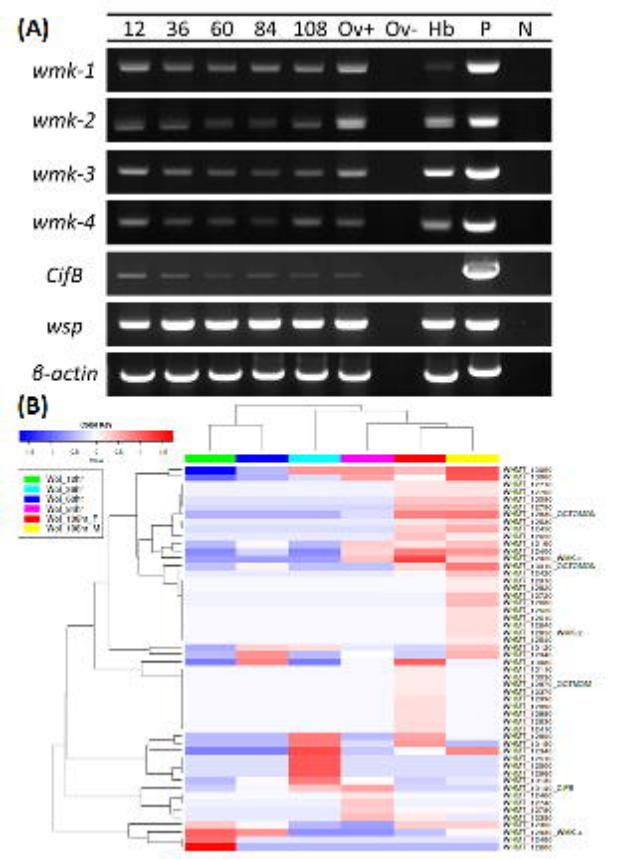


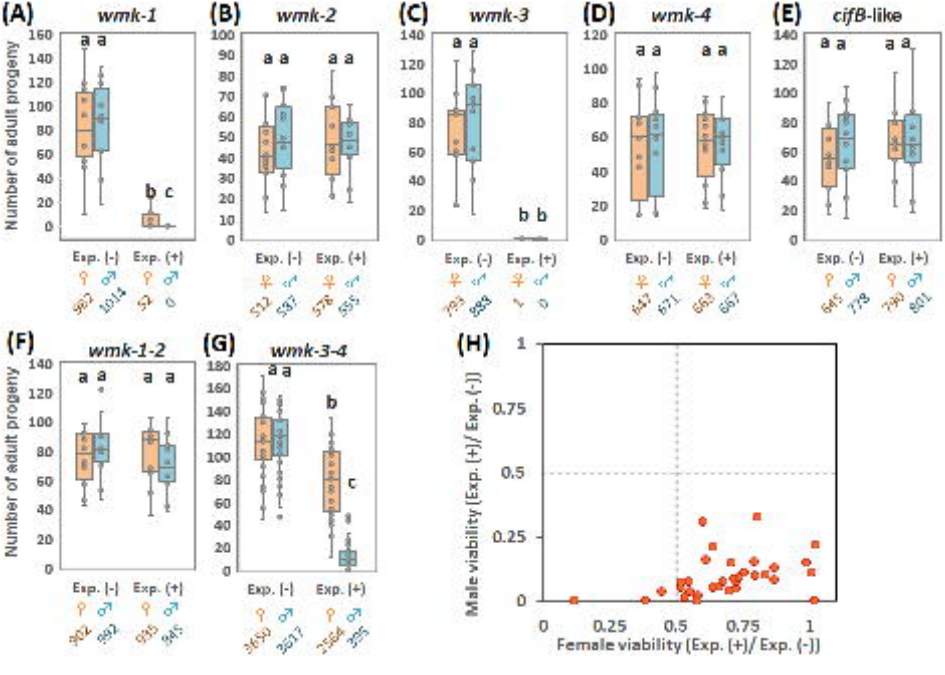
Protein cluster	Annotation	Gene 1	Gene 2	Gene 3	Gene 4
wHm-t specific Clus.1	Ankyrin repeat containing protein pk3 homolog	wHmt_09860	wHmt_12470		
wHm-t specific Clus.2	Mobile element protein	wHmt_01700	wHmt_11230	wHmt_11240	wHmt_12310
wHm-t specific Clus.3	Phage tail proteins	wHmt_11300	wHmt_12540		
wHm-t specific Clus.4	Hypothetical protein	wHmt_11790	wHmt_14170		
wHm-t specific Clus.5	Hypothetical protein	wHmt_01320	wHmt_01330		
wHm-t specific Clus.6	RNA directed DNA polymerase	wHmt_2000	wHmt_14180		
wHm-t absent Clus.1	Hypothetical protein	wHma_06950	wHmb_08320	wHmc_01340	
wHm-t absent Clus.2	Mobile element protein	wHma_09860	wHmb_01470	wHmb_05980	wHmc_10090

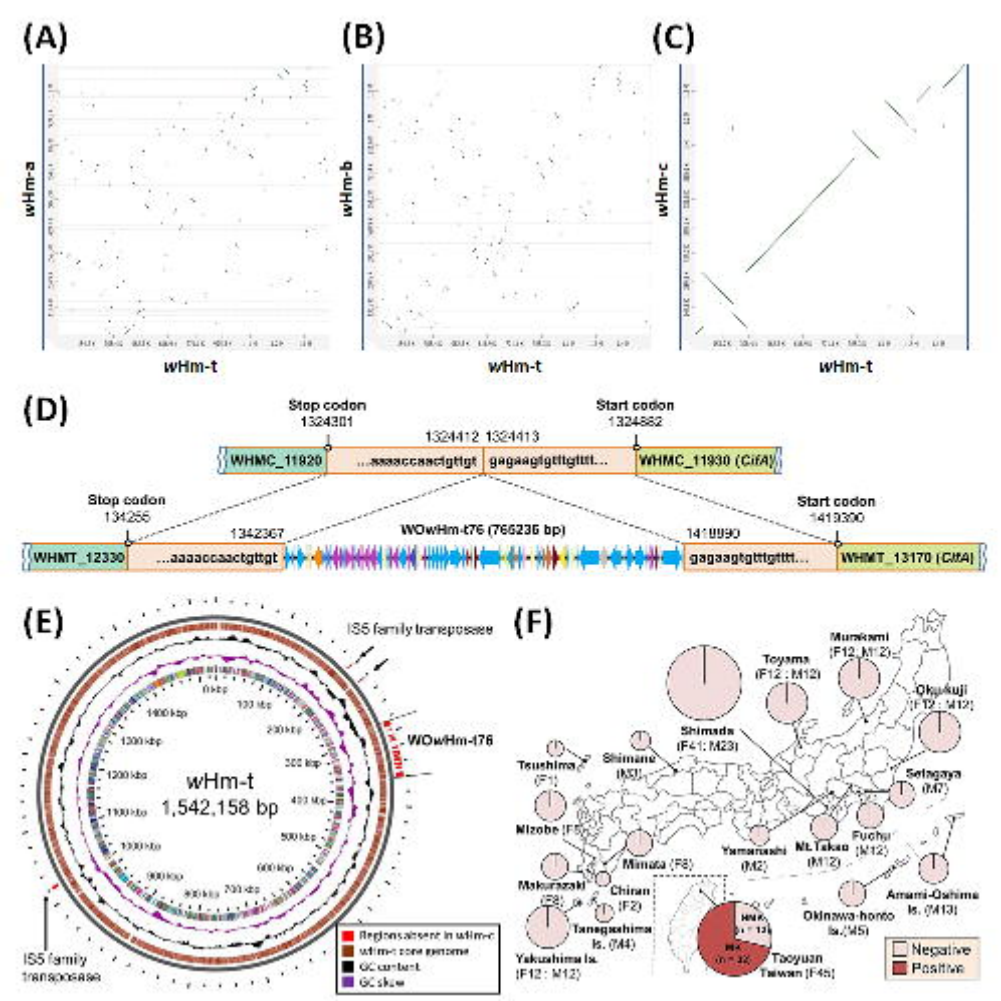

(A)
Region 7
Region 6
Region 5
Region 4
Region 3
Region 2
Region 1
wHm-t strain
Region 4
Region 1
Region 3
Region 2
wHm c strain
(B)
WOwHm-c76
Region 1
wHm-t
Region 2
wHm-c


(A)
(B)
(C)
II
III
I
0.05
Phage WO
Wolbachia
Homona
CifB-like
ORF-7
wsp
β-actin
Pur
W1/2
WOwKue (CI)
WOwHm-t76 (MK)
WOwCauB2 (CI)

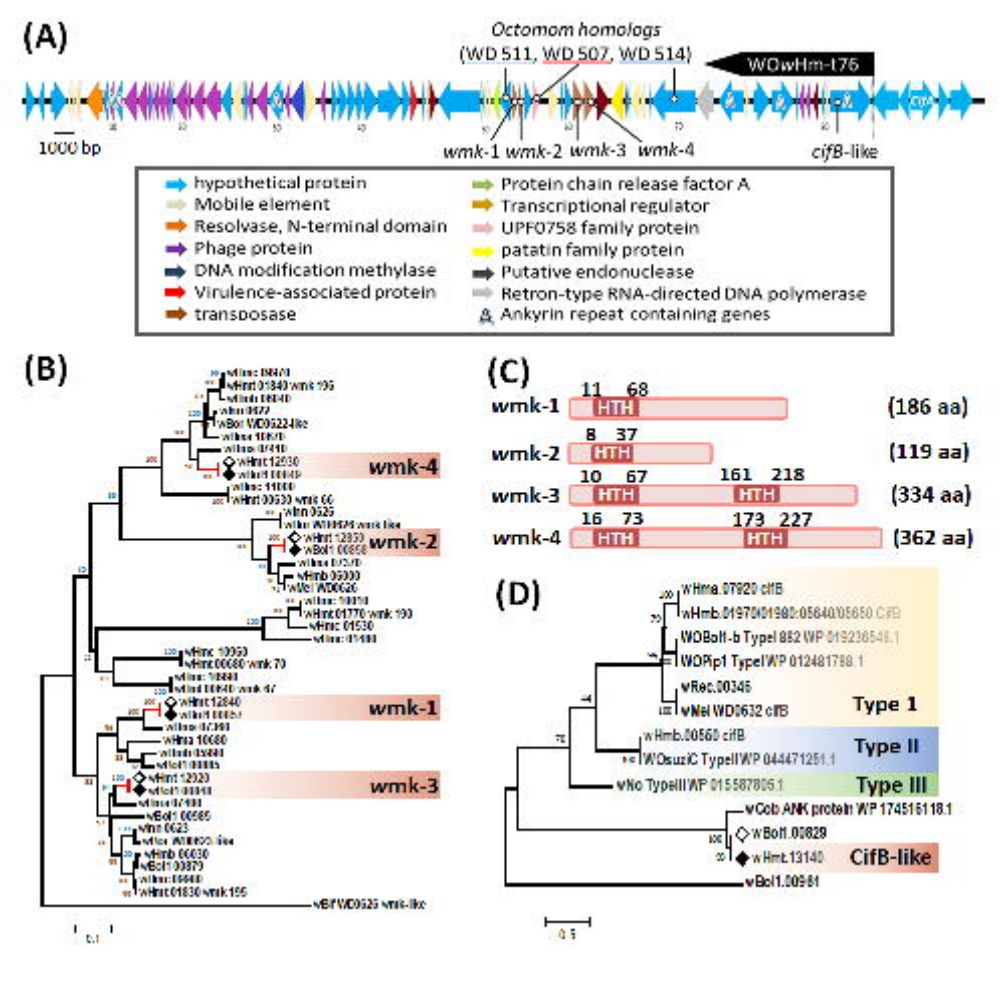

Wolbachia wsp copies (log10/10 ng DNA)
7
6
5
4
3
2
1
0
a
b
c
c
N = 128
N = 54
N = 4
N = 44
wHm-t MK-F
wHm-t NMK-F
wHm-c F
wHm-c M

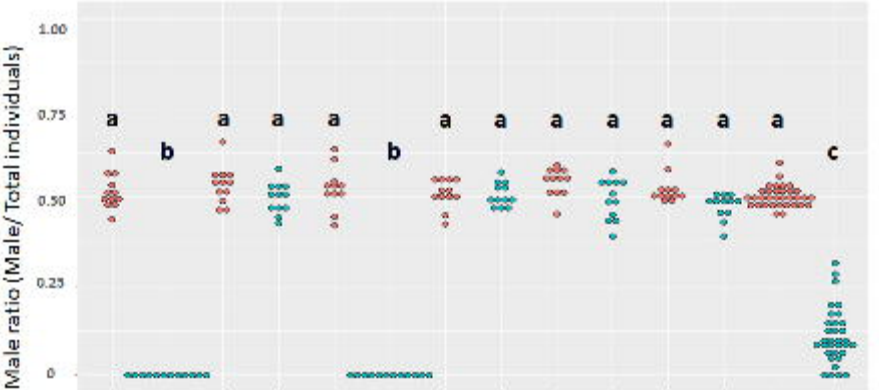

(A)
12
36
60
84
108
Ov+
Ov-
Hb
P
N
wmk-1
wmk-2
wmk-3
wmk-4
CifB
wsp
β-actin
(B)

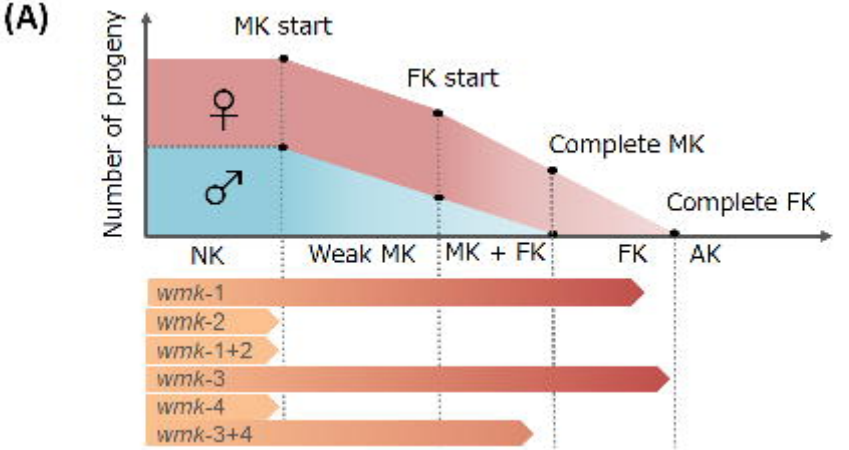

(A)
wmk-1
(B)
wmk-2
(C)
wmk-3
(D)
wmk-4
(E)
cifB-like
(F)
wmk-1-2
(G)
wmk-3-4
(H)
Number of adult progeny
Exp. (-)
Exp. (+)
Male viability (Exp. (+)/ Exp. (-))
Female viability (Exp. (+)/ Exp. (-))

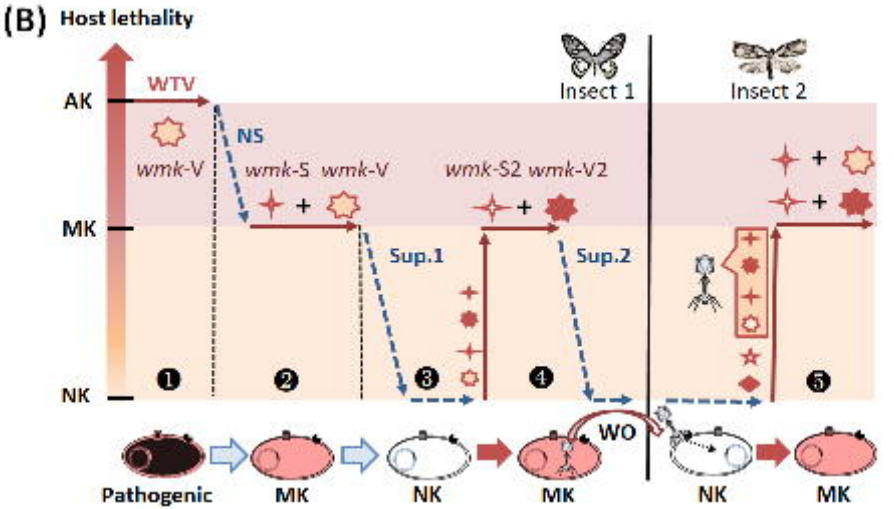


(A)
(B)
(C)
wHm-a
wHm-b
wHm-c
wHm-t
(D)
Stop codon
1324301
Start codon
1324882
1324412 1324413
WHMC_11920
...aaaaccaactgttgt
gagaagtgtttgtttt...
WHMC_11930 (CifA)
Stop codon
134255
Start codon
1419390
1342367
WOwHm-t76 (766236 bp)
1419990
WHMT_12330
...aaaaccaactgttgt
gagaagtgtttgtttt...
WHMT_13170 (CifA)
(E)
IS5 family transposase
WOwHm-t76
wHm-t
1,542,158 bp
IS5 family transposase
Regions absent in wHm-c
wHm-c core genome
GC content
GC skew
(F)
Murakami (F12 : M12)
Toyama (F12 : M12)
Okukuji (F12 : M12)
Shimada (F41 : M23)
Tsushima (F1)
Shimane (M3)
Setagaya (M7)
Nizobe (F5)
Yamanashi (M2)
Mt.Takao (M12)
Fuchu (M12)
Mimata (F8)
Makurazaki (F8)
Chinan (F2)
Amami-Oshima Is. (M13)
Okinawa-honto Is. (M5)
Tanegashima Is. (M4)
Yakushima Is. (F12 : M12)
Taoyuan Taiwan (F45)
Negative
Positive


(A)
Octomom homologs
(WD 511, WD 507, WD 514)
WOwHm-t76
1000 bp
wmk-1
wmk-2
wmk-3
wmk-4
cifB-like
hypothetical protein
Mobile element
Resolvase, N-terminal domain
Phage protein
DNA modification methylase
Virulence-associated protein
transposase
Protein chain release factor A
Transcriptional regulator
UPF0758 family protein
patatin family protein
Putative endonuclease
Retron-type RNA-directed DNA polymerase
Ankyrin repeat containing genes
(B)
wmk-4
wmk-2
wmk-1
wmk-3
(C)
wmk-1
11 68
HTH
(186 aa)
wmk-2
8 37
HTH
(119 aa)
wmk-3
10 67
161 218
HTH
HTH
(334 aa)
wmk-4
16 73
173 227
HTH
HTH
(362 aa)
(D)
Type 1
Type II
Type III
CifB-like


Male ratio (Male/ Total individuals)
1.00
0.75
0.50
0.25
0
a
b
a
a
a
b
a
a
a
a
a
a
a
c
wmk-1
wmk-2
wmk-3
wmk-4
cifB-like
wmk-1-2
wmk-3-4
Exp. (-) Exp. (+) Exp. (-) Exp. (+) Exp. (-) Exp. (+) Exp. (-) Exp. (+) Exp. (-) Exp. (+) Exp. (-) Exp. (+) Exp. (-) Exp. (+)
Fem. 982 52 512 578 793 1 647 663 645 790 902 935 3650 2564
Mal. 1014 0 587 555 888 0 671 667 778 801 992 845 3617 395


(A)
Number of progeny
MK start
FK start
Complete MK
Complete FK
♀
♂
NK
Weak MK
MK + FK
FK
AK
wmk-1
wmk-2
wmk-1+2
wmk-3
wmk-4
wmk-3+4


(B)
Host lethality
AK
MK
NK
WTV
NS
wmk-V
wmk-S
wmk-V
wmk-S2
wmk-V2
Sup.1
Sup.2
Insect 1
Insect 2
WO
Pathogenic
MK
NK
MK
NK
MK