

Evolution of Wolbachia reproductive and nutritional mutualism: insight from two novel strains that double infect pollinator of dioecious *Ficus hirta*

Wanzhen Liu

The Chinese Academy of Sciences

Xue Xia

Jiangsu Academy of Agricultural Sciences

Yamei Ding

The Chinese Academy of Sciences

Ji-Chao Fang

Jiangsu Academy of Agricultural Sciences

Hui Yu (✉ yuhui@scib.ac.cn)

The Chinese Academy of Sciences

Research Article

Keywords: Wolbachia-pollinator-host interactions, mutualism, cytoplasmic incompatibility, biotin, genome

Posted Date: April 28th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-2804106/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Genomics on November 1st, 2023. See the published version at <https://doi.org/10.1186/s12864-023-09726-2>.

Abstract

Wolbachia is a genus of endosymbiont which can manipulate the reproductive system of hosts and exert an influence on metabolic processes. *Wolbachia* is dominant in male syconium of dioecious *Ficus hirta* with pollinator *Valisia javana*. From high-quality chromosome-level *V. javana* genome with PacBio long-read and Illumina short read sequencing, we discovered a sizeable proportion of *Wolbachia* sequences and then assembled two novel strains belonging to supergroup A. Besides the basic genome description, we identified mobile genetic elements including prophages and insertion sequences, genes related to biotin synthesis and metabolism, and cytoplasmic incompatibility related *cif* genes in them. The results showed that (i) A total of two prophages and 201 insertion sequences were found. The prophage WOjav1 is cryptic and WOjav2 relatively intact. IS5 is dominant family. (ii) At least three pairs of *cif* genes belong to type I and is speculated to cause strong CI with infected female biased reproductive advantages; (iii) The *Wolbachia* only provide biotin precursors to absorb and metabolize biotin which is synthesized by host plant. This research provided a basis for reproductive and nutritional mutualism between *Wolbachia* and fig-fig wasp system, and will serve as a useful resource for further research into *Wolbachia*-pollinator-host interactions.

Background

Wolbachia is a genus of intracellular endosymbiotic bacteria belonging to the order Rickettsiales and is estimated to be distributed in ca. 44% arthropods, 66% insects, and also shows mutualistic symbiosis in nematodes [1–3]. *Wolbachia* is classified as a single species, *Wolbachia pipiensis*, divided into supergroups A to U, with supergroups A and B infecting arthropods exclusively based on the phylogenetic analysis using the 16S rDNA, *ftsZ* and *wsp* markers [4–6]. *Wolbachia* is transmitted mainly via vertical transmission from mother to offspring with high fidelity, but also can be transmitted across different taxa by host shift (HS), especially for supergroups A and B occurring between both phylogenetically close and more distant host species [4, 7, 8].

To increase the transmission and propagation, *Wolbachia* can manipulate the reproductive system of eukaryotic hosts using diverse methods, such as cytoplasmic incompatibility (CI), feminization of genetic males, parthenogenesis, and male killing [9–12]. CI is the most common reproductive manipulation effect induced by *Wolbachia*, and can be used as an important means to control arboviruses [13]. Co-expression of the prophage WO gene *cifA* and *cifB* in *Wolbachia* can cause a CI phenotype, which can be rescued by maternal *cifA* expression independently [14]. Phylogenetic analysis of *cif* protein encoded by *cif* gene shows that the evolution of *cifA* and *cifB* is consistent. Both *cifA* and *cifB* can be divided into five clades as type I–V, respectively [15–17].

Wolbachia can also exert an influence on numerous processes in the host, including immune, behavioral, and metabolic processes [18], which may be related to their extensive occupation of host cells and reproductive tissues [19]. It has been shown to be beneficial in insects by protecting them from pathogenic viral infection and increasing host fitness and fecundity [20–23]. In nematodes and *Cimex lectularius*, *Wolbachia* functions as an obligatory mutualistic endosymbiont [17, 24]. Therefore, *Wolbachia* has great research potential as well as application value.

Fig and fig wasps form a classical model in the study of coevolution and cospeciation [25]. Figs are plants in the genus *Ficus* (Moraceae), with around 750 species worldwide, distributed in the tropics and subtropics [26, 27]. Figs and fig-pollinating wasps (Agaonidae; Chalcidoidea; Hymenoptera) are obligate mutualists that have coevolved for over ~75 million years [28, 29]. The inflorescence of fig is an almost completely closed syconium, and has only one narrow ostiole communicated with the exterior. For the most part, only specialized fig wasps can enter the syconium, pollinate the seed flowers and lay eggs into some of their ovaries which called gall flowers at the same time. After a certain time, adults emerged and mated, then the female fig wasps fly out of the syconium with pollen in search of new receptive syconium to begin a new life cycle, while the flightless males of most species spend their entire lives within their natal syconium [30]. Except pollinators, there are many non-pollinating fig wasps in chalcidoidea from two families of the superfamily Chalcidoidea, including Epichrysomallinae family and subfamilies Sycophaginae and Pteromalinae (especially Otiteselini tribe with Sycoecina, Otiteselina and Sycoryctina subtribes) of Pteromalidae family, which feed on fig tissue or other wasps [31–33] and some of them lay eggs outside the syconium with remarkable long ovipositors. Different wasp species can survive within the same syconium. In most fig species, each syconium contains 50–500 wasps of a few different species developing together in close proximity [34]. As such, the system is an ideal case study in the role of ecological interactions in *Wolbachia* transmission and persistence.

More than half of the pollinating fig wasp species that were collected from North America, Australia, Asia and Africa have been found infected with *Wolbachia* [8, 34–37], which is remarkably higher than that in other insects [37]. The comparison of the endosymbionts diversity between female and male inflorescence in the dioecious *F. hirta* using high-throughput sequencing and biological databases showed *Wolbachia* is dominant in male syconium instead of the female, since that the later has no galled flowers for wasp larva [38]. Moreover, genomes of *Wolbachia* from three pollinating fig wasp species (*Ceratosolen solmsi*, *Kradibia gibbosae* and *Wiebesia pumilae*) have been characterized like were classified into supergroup A and found only one cryptic WO prophage in all of them, which is counterintuitive that *Wolbachia* with cryptic

prophages usually possesses at least one intact WO prophage consisting of gene sequences of the head, baseplate, and tail modules, through which the prophage could form intact virions [39].

A high-quality chromosome-level genome for *V. javana* Mayr (Agaonidae, Chalcidoidea, Hymenoptera), formerly known as *Blastophaga javana* Mayr, has been assembled using a combination of PacBio long-read sequencing and Illumina short read sequencing [40]. From *V. javana* genomics, we discovered a sizeable proportion of *Wolbachia* genomic sequences and then assembled two novel strains that are different from published *Wolbachia*. In this study, besides the basic genome description of these two strains of *Wolbachia*, we identified mobile genetic elements (MGEs) including prophages and insertion sequences, genes related to biotin (vitamin B7) synthesis and metabolism, and *cif* homologous genes in them. This genome assembly will serve as a useful resource for further research into *Wolbachia*-pollinator–host interactions, comparative genomics and coevolution.

Results

The high-coverage symbiont *Wolbachia* genome data include in the low-coverage *V. javana*

genome third-generation sequencing raw data

We found that there were a large number of *Wolbachia* genomic fragments in the sequence data of *V. javana* [40]. There were 861,951 reads in the sequence of *V. javana*, with a total length of 12,005.70 Mb and a coverage of about 40× to *V. javana* genome (296.34 Mb). After using BLASR to set different identities and maxscore, different numbers of *Wolbachia* reads were obtained (Figure 1A and 1B). We performed genome assembly using reads under different parameters, and the results showed that there were many sequences that did not belong to *Wolbachia* between the parameters maxscore -250 and -300 of the BLASR screen (Table 1). Using the BLASR parameters as Identity $\geq 60\%$ and maxscore ≤ -250 , the aligned reads were assembled, and the final genome size exceeded 50 Mb, which was seriously inconsistent with the normal size of the *Wolbachia* genome (Table S1). Therefore, we finally used the BLASR parameters Identity $\geq 60\%$ and maxscore ≤ -300 to assembling the third-generation sequencing data of about 610 Mb. According to the assembled *Wolbachia* longest genome of 2,352,936 bp, the coverage is more than 270×.

***V. javana* double infection with two novel *Wolbachia* strains**

According to the length of the assembled genome and the number of fragments, we selected the sequence of 2,352,926 bp (2.24 Mb) as the final genomic sequence (Table 1). After polished by the *Wolbachia* second-generation sequencing data, a genomic sequence with a length of 2,352,827 bp was obtained. The GC content of the genome was 35.0%, CDS and RNAs were 2,825 and 62, respectively (Table 2). This genome size is much larger than the known *Wolbachia* (Table S1). Therefore, the MLST sequences from the *wMel* strain infected *Drosophila melanogaster*, whose number is ST = 1, were used to search for five MLST housekeeping genes in *Wolbachia*. There were at least two sets of MLST sequences in the genome (Table 3). These MLST sequences do not belong to any known *Wolbachia* strain, so they are two novel strains, and referred to as *wJav1* and *wJav2*.

Identification of *wJav1* and *wJav2*

Since the *Wolbachia* genomes double infected host *V. javana* and difficult to separate, we used different combinations of MLST sequences for the identification. According to allele of five housekeeping genes, the MLST sequences in *wJav1* and *wJav2* can be divided into two groups (Table 4), and compared with the known MLST sequences in the database for each combination. Difference between *wJav1* and *wJav2* of two potential groups are two types of allelic profiles of *gatB* and *coxA*, separately (Table 4). *wJav2* of group 1 is the most similar to the strain Sole_A_17-02 infected fire ant (*Solenopsis*, Apocrita, Hymenoptera) which belong to supergroup A (Figure 2, Table S2). Only allele of the *hcpA* gene was different between Sole_A_17-02 and *wJav2* strains (Table 4), which indicated that these two strains may be obtained by the same strain after horizontal transmission among different hosts. *gatB* and *coxA* are divided into two types for *wJav*, respectively (Figure 2). According to the unique *hcpA* = 60 and *ftz* = 3 of *V. javana Wolbachia*, two most similar strains were found in the PubMLST. Their hosts were *Sogatella furcifera* (Sfur_A_YN4, Hemiptera, Delphacidae) and *Megastigmus* sp.3 (Msp_i_A_wspe, Hymenoptera, Torymidae). The former is a major pest in rice paddies throughout Asia [41], and the latter are serious pests of conifer species with four basic feeding types [42]. We randomly selected 21 *Wolbachia* MLST sequences (Table S2) belong to different supergroups (A, B, D, F, H) occurred in different hosts to construct the evolutionary tree by maximum likelihood (ML). The ML tree showed that both *wJav1* and *wJav2* belong to supergroup A and are closely related to each other (Figure 2).

The relationship between the mixed genome and other 313 *Wolbachia* genomes was further evaluated according to the average nucleotide identity (ANI) (Table S3). The mixed genome has the highest ANI with the *Wolbachia* sequences belonging to supergroup A (Table 5). Combined ANI result with MLST ML tree (Figure 2), it could be sure that both *wJav1* and *wJav2* belong to supergroup A.

Mobile genetic elements (MGEs) in *wJav1* and *wJav2*

MGEs is an important source for bacteria to acquire foreign genes and is an important reason for genome recombination. MGEs mainly include prophage and insertion sequence (IS) in bacteria. In the mixed genome of *wJav1* and *wJav2*, a total of two prophages (Figure 3B) and 201 insertion sequences (Figure 3A) were found. The main types of ISs are similar to supergroup A *Wolbachia* with IS5, IS4, IS630 and IS110 families accounting for 38.28%, 19.92%, 16.8% and 12.5%, respectively. As dominant IS family, there are two subgroups of IS5 including IS5_ssgr_IS903 (n = 43) and IS5_ssgr_IS1031 (n = 21). The IS families are consistent with fig wasp *Ceratosolen solmsi* but numbers, in which there are 201 IS of the *V. javana* and 86 IS for *C. solmsi* [43]. Two prophages were referred to as WOjav1-2, and most of the regions in them (Figure 3B) are phage-like proteins, hypothetical proteins and transposase proteins (Figure 4; Table S5). WOjav1 is cryptic and only contains transposase protein and minor tail protein. WOjav2 is relatively intact, with complete phage replication and assembly related genes, such as transposase, head morphogenesis protein, baseplate assembly protein with lysozyme, and tail protein. (Table S5).

Cif genes in *wJav1* and *wJav2*

In total, 4 genes encoded proteins were found to be aligned with different types (I and IV) of CifA and 57 genes encoded proteins were found to be aligned with different types (I, II and V) of CifB in the *wJav1* and *wJav2* genomes (Table S6). *cifA* and *cifB* genes transcribed together can manipulate host insect reproduction by cytoplasmic incompatibility (CI) [16]. Fig|953.597.peg. 1691 (335 bp) and fig | 953.597.peg 1692 (1112 bp) may be the same *cifA* gene, and it is annotated as two genes because it contains multiple stop codons (Table S6). So subsequent phylogenetic analysis will analyze the two genes in tandem as one gene. Fig | 953.597.peg 384 (1478 bp) and fig | 953.597.peg 1839 (1511 bp) are the *cifA* genes (Table S6). The alignment length between fig|953.597.peg.995 and type V *cifB* is too short, and the alignment position is 3703aa - 3939aa (Table S6). This part is an ANK protein domain, not a key conserved region of the CifB, so it will not be analyzed. Therefore, at least three different pairs of *cif* genes (*wJav_pair1-3* for short) are included in *wJav1* and *wJav2* genomes (Table S6). The ML tree constructed by *wJav_pair1-3* and other different types of *cif* genes (Table S4 and Table S7 [17]) indicated that all *wJav_pair1-3* belong to type I, and *wJav_pair1-2* were very close which may be differentiated from one same gene (Figure 5).

Genes related to biotin synthesis and metabolism in *wJav1* and *wJav2*

According to KEGG pathway enrichment, the *wJav1* and *wJav2* mixed genome do not contain any key genes (*BioF/A/D/B*) for biotin synthesis. However, the genes of glutaryl-[acp] methyl ester and pimeloyl-[acyl carrier protein] methyl ester, *fabF/G/Z/I*, which are the key precursors for the synthesis of biotin, and the genes *birA*, which metabolize biotinyl-5'-AMP and holo-[carboxylase] in the downstream, are included (Figure 6). The genome-wide functional annotation based on the RAST and eggnoG web servers also proved that the mixed genome do not contain the key genes for the synthesis of biotin, but include the gene *bioY* to uptake biotin from the host, and *birA* to synthesis their own metabolites using biotin as an essential precursor (Table 6). So, we can infer that *wJav1* and *wJav2* do not directly provide biotin for the host, but can provide biotin precursors, and interact with other microorganisms (such as *Cardinium*) with complete biotin synthesis pathway to provide biotin for the host. At the same time, *wJav1* and *wJav2* have the ability to absorb and metabolize essential biotin from the host.

Discussion

The high proportion sequencing data of *Wolbachia* in the third-generation sequencing data of *V. javana* indicates its high titer and potential important role in the pollinator. *Wolbachia* could not proliferate without restraint in the host. The host will evolve some mechanisms to limit the titer of *Wolbachia* in the host. For example, the maternal effect gene *Wolbachia identity suppressor* (*Wds*) of *Nasonia* can control the titer of *Wolbachia* in the body and inhibit the adverse effects caused by *Wolbachia* over-amplification [44]. In this study, the *Wolbachia* genome with high coverage was found in the low coverage *V. javana* genome data from the third-generation sequencing, which may indicate that *Wolbachia* in *V. javana* was not inhibited by the host or highly proliferated due to its beneficial effect on the host.

The pollinator *V. Javana* was doubly infected by two strains of *Wolbachia*, both of strains belonged to supergroup A and had closed relationship (Fig. 2, Table S2, Table S3). This result is basically consistent with previous studies, that is, fig wasps could be infected by multiple strains of *Wolbachia*, and the strains infected pollinators are mainly supergroup A. Gene markers such as *wsp* [8, 34, 45] were applied for analysis in first, MLST [46] and genome data [39] come later. The double infection situations of *Wolbachia* could be divided into (i) Different supergroups [47, 48], (ii) Different strains of a same supergroup [34], and (iii) Different strains of different main key gene operons [49]. And the possible causes of double infection include: (i) *Wolbachia* has been introduced into the host many times, (ii) *Wolbachia* began to diverge after invasion and transferred horizontally between host species, and (iii) The common ancestor was multiple infected, and then it is lost or acquired in the process of divergence [50]. The synergism between different *Wolbachia* strains may be beneficial to maintain the double infection. The double infection of the host produces a new compatible type, which may promote its transmission in the host population, and also creates a place for the gene exchange between bacterial strains [48]. In addition, it increases the complexity of the phylogenetic characteristics of a single strain, and the determination of the relative contribution of each strain to the overall density [51].

The existence of multiple pairs of *cif* genes indicates that *wJav1* and *wJav2* are highly likely to induce strong CI. The MLST result showed that related species of *V. javana* infected with *Wolbachia* and parts of them cause CI, however, it seems will not be limited by phylogeny totally

(Fig. 2). The endosymbiont of the most related species *Dactylopius coccus* belongs to type I (Fig. 5, Table S7) and will lead to CI [52]. Double infected host will get CI mainly [48]. It has been reported that the *cif* gene with Type I can cause and save CI through transgenic *Drosophila* [53]. *Wolbachia* CI mode is related to *cifA/cifB* homolog similarity and copy number, and different strains have different CI intensity [15]. A strain with two or three copies induce strong CI, such as *wRi* and *wHa* [15]. Our results showed that the *cifA/cifB* of *F. hirta* pollinator is type I with a copy number of three, which is a branch with *wRi* and *wHa*, so it is speculated that it has a strong CI strength.

From the perspective of CI, the unique life history and mating behavior of pollinator fig wasps promote the spread of *Wolbachia*. *F. hirta* is a dioecious plant, of which the number of female trees in the field is far greater than that of male trees [54]. The whole life cycle of the male *V. javana* only occurs in the figs of male trees (gall figs). The female wasp has completed mating before flying out of gall figs, while the wasp only spawn in the gall flowers in gall figs [55]. Therefore, as long as the wasp is infected by *Wolbachia*, the infection will rapidly spread in the local population under the action of CI. In short, the infected male *V. javana* without wing could not leave the male trees, while the infected female pollinator can fly with the wind or take the initiative to spread *Wolbachia* within a certain range [56–58]. The male trees of the *V. javana* infected with *Wolbachia* will make the offspring of the non-infected females sterile, thus making the infected females obtain stronger reproductive advantages. When the *Wolbachia*-uninfected female *V. javana* enters gall fig that have *Wolbachia*-infected male pollinator to lay eggs, because the offspring of the *Wolbachia*-uninfected female *V. javana* must not be uninfected at this time, and the offspring produced after mating with the infected male cannot be hatched. Further from the perspective of long-term evolution, the fig wasps living on each male tree are regarded as a small population, and the male trees have *Wolbachia*-infected *V. javana* in the early stage will be more likely to transmit male pollen due to the reproductive advantage of *Wolbachia* infection, which is conducive to the preservation and transmission of the genes of the male tree itself.

Wolbachia will helps the pollinator to absorb and metabolize biotin which is synthesized by host plant. Biotin is an essential vitamin for insects [59]. It functions as a covalently-bound cofactor in various carboxylases, that have major roles in fatty acid biosynthesis, amino acid and fatty acid catabolism, and the citric acid cycle [60, 61]. To figure out the distribution of biotin-related genes in *F. hirta*, genome data (bract [62], male and female flowers [63], male and female flowers: unpublished), pollinator *V. javana* and its symbiosis *Wolbachia* were compared. We found that *F. hirta* has biotin-synthesis genes and the bract of the fig can transcribe most biotin-related genes (Table S8). There are fewer genes of female and male flowers of male fig tree related to biotin, while female flowers have *bioA*, *pdhC* and *sucB* compared to male flowers (Table S8). The *V. javana* also have biotin related genes for absorb and metabolize, but fewer genes compared to the endosymbiont (Table S8). Consider endosymbiont embedded in the insect genome, related genes of pollinators may transfer from *Wolbachia*, for example, through horizontal gene transfer genes [64]. The pollinator lays its eggs in the gall flower, and its larvae feed on the gall flowers where there are biotin synthesis genes. Therefore, it is speculated that the biotin of pollinators acquired from the *F. hirta* gall flowers, and *Wolbachia* helps to absorb and metabolize biotin. Due to feed by the food that unbalanced nutrition or lack Vitamin B, *Wolbachia* of *Bemisia tabaci* complex [65], Planthoppers [66] and exclusive blood feeders maintain the function of synthesizing biotin, and even has streamlined biotin operons. However, *Wolbachia* of *V. javana* is different, it is more likely that the *F. hirta* provides enough nutritionally balanced food and does not need the assistance of endosymbiont, it might reflect that after a long evolution history, synergetic relationship between pollinators, endosymbiont and *F. hirta* has become closer. And the related genes of the pollinator might help to understand that why some uninfected offspring could survive without reproductive and nutritional advantage.

Conclusions

We have provided the assembly of a mixed genome infected *V. javana* by two novel *Wolbachia* strains belonging to supergroup A. Basic genomic analyses revealed genome features, mobile genetic elements (MGEs) including prophage and insertion sequences, genes related to biotin synthesis and metabolism, and *cif* homologous genes were found in them. The MLST obtained from the mixed genome *wJav1-2* will further understand the phylogenetic relationship between different *Wolbachia* supergroups, and conduct further biochemical, molecular and genetic analysis of *wJav1-2* and related *Wolbachia*. The annotation and analysis of prophages and mobile elements highlight their significant impact on genome evolution and gene content in intracellular symbiosis, indicating that these elements can be reused as gene manipulation tools for *Wolbachia*. The discovery of the *cif* genes further confirmed the existence of CI in the pollinating fig wasp of *F. hirta*, and provided a basis for further research on the co-evolution of CI among symbiotic bacteria, insects and plants. Based on the understanding of biotin pathway and related genes of *F. hirta* - pollinators - endosymbiont, it provides ideas and data for further exploring its highly related nutritional symbiosis relationship. This genome assembly will serve as a useful resource for further research into *Wolbachia*-pollinator–host interactions, comparative genomics and coevolution.

Methods

Genomes assembly of *Wolbachia*

We collected nearly 500–1,000 female adult samples of *V. javana* from several figs of *F. hirta* in a single tree in Baiyun Mountain, Guangdong Province of China. The *Wolbachia* genomes were assembled based on the whole genome sequencing data (accession number of

GWHBDGE000000000 for *V. javana*).

In this study, 319 assembled versions of *Wolbachia* genomes including 154 strains from different host species downloaded by NCBI were used as the reference genome (Table S1). The third-generation sequencing raw data of *V. javana* genome (coverage 40 ×; 11.9 GB) were aligned to the *Wolbachia* reference genome using BLASR [67] to obtain the reads of the symbiotic *Wolbachia*. The second-generation sequencing reads of *Wolbachia* were acquired by aligning the second-generation genome sequences (Illumina data) of *V. javana* were aligned to the *Wolbachia* reference genome using kneaddata (<https://huttenhower.sph.harvard.edu/kneaddata>). Then *Wolbachia* third-generation sequencing reads in *V. javana* were assembled using Flye assembler v2.7.1 [68], and the result of assembly was polished in Pilon version 1.23 [69] with Illumina data.

Genome annotation

The genome size and GC content of the assembled *Wolbachia* were calculated with seqkit v0.9.2 [70]. Rapid genome annotation was conducted by RAST server (<https://rast.nmpdr.org/rast.cgi>). Insert sequences (ISs) were founded in the IS database using ISSaga web server http://issaga.biotoul.fr/issaga_index.php [71]. The prophage regions in *Wolbachia* genome were annotated using Phaster web server (<http://phaster.ca>), and the visualization of the positions of the prophage regions was automatically generated by the Phaster web server. Functions and potential involved pathways of the coding genes in *Wolbachia* genome were further annotated using eggNOG Mapper web server (<http://eggno-mapper.embl.de>) [72] and the KEGG automatic annotation server KAAS (<https://www.genome.jp/tools/kaas/>) [73].

Identification and phylogenetic analysis of *Wolbachia* in *V. javana*

To identify the *Wolbachia* strains in *V. javana*, five multilocus sequence typing genes (MLST) including *gatB*, *fbpA*, *coxA*, *hcpA* and *ftsZ* in genomic data were blasted by BLASTN v2.9.0 according to the reference wMel strain of *Drosophila melanogaster* whose sequence type is typical ST1 (Table S2). The sequence type of the *Wolbachia* in *V. javana* were identified by alignment with PubMLST *Wolbachia* MLST database (<http://www.pubmlst.org/Wolbachia/>). In addition, ST sequences (sequence synthesized by housekeeping genes) of MLST in other *Wolbachia* genomes of A, B, D, F and H supergroups were downloaded from PubMLST *Wolbachia* MLST database (Table S2).

Multiple sequence alignment for these MLST genes were aligned by MAFFT v7.310 [74]. The most suitable substitution model was calculated by ModelFinder [75], and the maximum-likelihood (ML) tree was constructed by IQ-tree v1.6.10 [76]. In addition, the average nucleotide similarity (ANI) between *Wolbachia* genomes was calculated using fastANI to further quantify the relationship between this genome and other *Wolbachia* genomes (Table S3) [77].

Identification of *cif* genes

Based on the typical CifA and CifB protein sequences of type I-V (Table S4), the *cif* genes and its homologous in *Wolbachia* of *V. javana* were blasted by blastp v2.9.0. The selection of homologues was based on Lindsey's method [16] and slightly improved by the following criteria: 1) $e \leq 10^{-30}$; 2) percentage of identity matches $\geq 50\%$. The obtained *cif* genes were used to construct the ML tree with the *cif* genes from other *Wolbachia* strains [17].

Declarations

Acknowledgments

We thank Songle Fan, South China Botanical Garden, for his knowledge sharing on fig and fig wasp ecological and molecular characteristics. We also thank Weicheng Huang, South China Botanical Garden, provided annotation information on *Ficus hirta*.

Authors' contributions

H.Y. supervised the project. X. X. conducted data analysis. All authors contributed to writing the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (Grants Nos. 31971568), the Chinese Academy of Sciences PIFI Fellowship for Visiting Scientists (2022VBA0002) and Guangdong Basic and Applied Basic Research Foundation (2021A1515110981).

Availability of data and materials

Data for this study will be completed after manuscript is accepted for publication. All genome sequence data of *Wolbachia* are available in GWH database of the National Genomics Data Center (NGDC) with accession no. GWHCBHN000000000. The raw genome sequencing data of Illumina and PacBio have deposited in the GSA database of the National Genomics Data Center (NGDC) with accession number of CRA010785.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Werren JH: **Biology of Wolbachia**. *Annu Rev Entomol* 1997, **42**:587-609.
2. Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH: **How many species are infected with Wolbachia?—a statistical analysis of current data**. *Fems Microbiol Lett* 2008, **281**(2):215-220.
3. 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* 2012, **7**(6).
4. Gomes TMFF, Wallau GL, Loreto ELS: **Multiple long-range host shifts of major Wolbachia supergroups infecting arthropods**. *Sci Rep-Uk* 2022, **12**(1).
5. Laidoudi Y, Levasseur A, Medkour H, Maaloum M, Ben Khedher M, Sambou M, Bassene H, Davoust B, Fenollar F, Raoult D *et al*: **An Earliest Endosymbiont, Wolbachia massiliensis sp. nov., Strain PL13 from the Bed Bug (Cimex hemipterus), Type Strain of a New Supergroup T**. *Int J Mol Sci* 2020, **21**(21).
6. Olanratmanee P, Baimai V, Ahantari A, Trinachartvanit W: **Novel supergroup U Wolbachia in bat mites of Thailand**. *Southeast Asian Journal of Tropical Medicine and Public Health* 2021, **52**(1):48-55.
7. Oliver KM, Degan PH, Burke GR, Moran NA: **Facultative Symbionts in Aphids and the Horizontal Transfer of Ecologically Important Traits**. *Annu Rev Entomol* 2010, **55**:247-266.
8. Shoemaker DD, Machado CA, Molbo D, Werren JH, Windsor DM, Herre EA: **The distribution of Wolbachia in fig wasps: correlations with host phylogeny, ecology and population structure**. *P Roy Soc B-Biol Sci* 2002, **269**(1506):2257-2267.
9. Hurst GDD, Hammarton TC, Bandi C, Majerus TMO, Bertrand D, Majerus MEN: **The diversity of inherited parasites of insects: the male-killing agent of the ladybird beetle *Coleomegilla maculata* is a member of the Flavobacteria**. *Genetics Research* 1997, **70**(1):1-6.
10. O'Neill SL, Giordano R, Colbert A, Karr TL, Robertson H: **16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects**. *Proceedings of the national academy of sciences* 1992, **89**(7):2699-2702.
11. Rousset F, Bouchon D, Pintureau B, Juchault P, Solignac M: **Wolbachia Endosymbionts Responsible for Various Alterations of Sexuality in Arthropods**. *P Roy Soc B-Biol Sci* 1992, **250**(1328):91-98.
12. Werren JH, Baldo L, Clark ME: **Wolbachia: master manipulators of invertebrate biology**. *Nat Rev Microbiol* 2008, **6**(10):741-751.
13. Flores HA, de Bruyne JT, O'Donnell TB, Nhu VT, Giang NT, Trang HTX, Van HTT, Long VT, Dui LT, Huy HL *et al*: **Multiple Wolbachia strains provide comparative levels of protection against dengue virus infection in *Aedes aegypti***. *Plos Pathog* 2020, **16**(4).
14. Shropshire JD, On J, Layton EM, Zhou H, Bordenstein SR: **One prophage WO gene rescues cytoplasmic incompatibility in *Drosophila melanogaster***. *P Natl Acad Sci USA* 2018, **115**(19):4987-4991.
15. LePage DP, Metcalf JA, Bordenstein SR, On JM, Perlmutter JI, Shropshire JD, Layton EM, Funkhouser-Jones LJ, Beckmann JF, Bordenstein SR: **Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility**. *Nature* 2017, **543**(7644):243-+.
16. Lindsey ARI, Rice DW, Bordenstein SR, Brooks AW, Bordenstein SR, Newton ILG: **Evolutionary Genetics of Cytoplasmic Incompatibility Genes *cifA* and *cifB* in Prophage WO of Wolbachia**. *Genome Biol Evol* 2018, **10**(2):434-451.
17. Martinez J, Klasson L, Welch JJ, Jiggins FM: **Life and Death of Selfish Genes: Comparative Genomics Reveals the Dynamic Evolution of Cytoplasmic Incompatibility**. *Mol Biol Evol* 2021, **38**(1):2-15.
18. Landmann F: **The Wolbachia Endosymbionts**. *Microbiol Spectr* 2019, **7**(2).
19. Duron O, Gottlieb Y: **Convergence of Nutritional Symbioses in Obligate Blood Feeders**. *Trends Parasitol* 2020, **36**(10):816-825.
20. Guo Y, Hoffmann AA, Xu XQ, Zhang X, Huang HJ, Ju JF, Gong JT, Hong XY: **Wolbachia-induced apoptosis associated with increased fecundity in *Laodelphax striatellus* (Hemiptera: Delphacidae)**. *Insect Mol Biol* 2018, **27**(6):796-807.
21. Moriyama M, Nikoh N, Hosokawa T, Fukatsu T: **Riboflavin Provisioning Underlies Wolbachia's Fitness Contribution to Its Insect Host**. *Mbio* 2015, **6**(6).

22. Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T: **Evolutionary origin of insect-Wolbachia nutritional mutualism.** *P Natl Acad Sci USA* 2014, **111**(28):10257-10262.
23. Zug R, Hammerstein P: **Bad guys turned nice? A critical assessment of Wolbachia mutualisms in arthropod hosts.** *Biol Rev* 2015, **90**(1):89-111.
24. Comandatore F, Sasser D, Montagna M, Kumar S, Koutsovoulos G, Thomas G, Repton C, Babayan SA, Gray N, Cordaux R *et al*: **Phylogenomics and Analysis of Shared Genes Suggest a Single Transition to Mutualism in Wolbachia of Nematodes.** *Genome Biol Evol* 2013, **5**(9):1668-1674.
25. Janzen DH: **How many parents do the wasps from a fig have?** *Biotropica* 1979:127-129.
26. Ronsted N, Weiblen GD, Clement WL, Zerega NJC, Savolainen V: **Reconstructing the phylogeny of figs (Ficus, Moraceae) to reveal the history of the fig pollination mutualism.** *Symbiosis* 2008, **45**(1-3):45-55.
27. Ronsted N, Weiblen GD, Savolainen V, Cook JM: **Phylogeny, biogeography, and ecology of Ficus section Malvanthera (Moraceae).** *Mol Phylogenet Evol* 2008, **48**(1):12-22.
28. Cruaud A, Ronsted N, Chantarasuwan B, Chou LS, Clement WL, Couloux A, Cousins B, Genson G, Harrison RD, Hanson PE *et al*: **An Extreme Case of Plant-Insect Codiversification: Figs and Fig-Pollinating Wasps.** *Syst Biol* 2012, **61**(6):1029-1047.
29. Machado CA, Jousset E, Kjellberg F, Compton SG, Herre EA: **Phylogenetic relationships, historical biogeography and character evolution of fig-pollinating wasps.** *P Roy Soc B-Biol Sci* 2001, **268**(1468):685-694.
30. Galil J, Eisikowitch D: **On the pollination ecology of Ficus sycomorus in East Africa.** *Ecology* 1968, **49**(2):259-269.
31. Rasplus J-Y, Kerdelhué C, Le Clainche I, Mondor G: **Molecular phylogeny of fig wasps Agaonidae are not monophyletic.** *Comptes Rendus de l'Académie des Sciences-Series III-Sciences de la Vie* 1998, **321**(6):517-527.
32. Campbell B, Heraty J, Rasplus J, Chan K, Steffen-Campbell J, Babcock C: **Molecular systematics of the Chalcidoidea using 28S-D2 rDNA.** *Hymenoptera evolution, biodiversity and biological control CSIRO Publishing, Collingwood, Australia* 2000:59-73.
33. Heraty JM, Burks RA, Cruaud A, Gibson GA, Liljeblad J, Munro J, Rasplus JY, Delvare G, Janšta P, Gumovsky A: **A phylogenetic analysis of the megadiverse Chalcidoidea (Hymenoptera).** *Cladistics* 2013, **29**(5):466-542.
34. Haine ER, Cook JM: **Convergent incidences of Wolbachia infection in fig wasp communities from two continents.** *P Roy Soc B-Biol Sci* 2005, **272**(1561):421-429.
35. Chen LL, Cook JM, Xiao H, Hu HY, Niu LM, Huang DW: **High incidences and similar patterns of Wolbachia infection in fig wasp communities from three different continents.** *Insect Sci* 2010, **17**(2):101-111.
36. Ahmed MZ, Greyvenstein OFC, Erasmus C, Welch JJ, Greeff JM: **Consistently high incidence of Wolbachia in global fig wasp communities.** *Ecol Entomol* 2013, **38**(2):147-154.
37. Wang NX, Jia SS, Xu H, Liu Y, Huang DW: **Multiple Horizontal Transfers of Bacteriophage WO and Host Wolbachia in Fig Wasps in a Closed Community.** *Front Microbiol* 2016, **7**.
38. Liu YF, Fan SL, Yu H: **Analysis of Ficus hirta Fig Endosymbionts Diversity and Species Composition.** *Diversity-Basel* 2021, **13**(12).
39. Miao YH, Xiao JH, Huang DW: **Distribution and Evolution of the Bacteriophage WO and Its Antagonism With Wolbachia.** *Front Microbiol* 2020, **11**.
40. Chen LF, Feng C, Wang R, Nong XJ, Deng XX, Chen XY, Yu H: **A chromosome-level genome assembly of the pollinating fig wasp Valisia javana.** *DNA Res* 2022, **29**(3).
41. Zhou C, Yang H, Wang Z, Long GY, Jin DC: **Comparative transcriptome analysis of Sogatella furcifera (Horvath) exposed to different insecticides.** *Sci Rep-Uk* 2018, **8**.
42. Roques A, Skrzypczynska M: **Seed-infesting chalcids of the genus Megastigmus Dalman, 1820 (Hymenoptera : Torymidae) native and introduced to the West Palearctic region: taxonomy, host specificity and distribution.** *J Nat Hist* 2003, **37**(2):127-238.
43. Neupane S, Bonilla SI, Manalo AM, Pelz-Stelinski KS: **Complete de novo assembly of Wolbachia endosymbiont of Diaphorina citri Kuwayama (Hemiptera: Liviidae) using long-read genome sequencing.** *Sci Rep-Uk* 2022, **12**(1):125.
44. Funkhouser-Jones LJ, van Opstal EJ, Sharma A, Bordenstein SR: **The Maternal Effect Gene Wds Controls Wolbachia Titer in Nasonia.** *Curr Biol* 2018, **28**(11):1692+.
45. Yang CY, Xiao JH, Niu LM, Ma GC, Cook JM, Bian SN, Fu YG, Huang DW: **Chaos of Wolbachia Sequences Inside the Compact Fig Syconia of Ficus benjamina (Ficus: Moraceae).** *Plos One* 2012, **7**(11).
46. Baldo L, Hotopp JCD, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, Hayashi C, Maiden MCJ, Tettelin H, Werren JH: **Multilocus sequence typing system for the endosymbiont Wolbachia pipientis.** *Appl Environ Microb* 2006, **72**(11):7098-7110.
47. van Meer MMM, Witteveldt J, Stouthamer R: **Phylogeny of the arthropod endosymbiont Wolbachia based on the wsp gene.** *Insect Mol Biol* 1999, **8**(3):399-408.

48. Werren JH, Zhang W, Guo LR: **Evolution and Phylogeny of Wolbachia - Reproductive Parasites of Arthropods.** *P Roy Soc B-Biol Sci* 1995, **261**(1360):55-63.
49. PerrotMinnot MJ, Guo LR, Werren JH: **Single and double infections with Wolbachia in the parasitic wasp *Nasonia vitripennis*: Effects on compatibility.** *Genetics* 1996, **143**(2):961-972.
50. Zhang YK, Zhang KJ, Sun JT, Yang XM, Ge C, Hong XY: **Diversity of Wolbachia in Natural Populations of Spider Mites (genus *Tetranychus*): Evidence for Complex Infection History and Disequilibrium Distribution.** *Microb Ecol* 2013, **65**(3):731-739.
51. Jeffries CL, Cansado-Utrilla C, Beavogui AH, Stica C, Lama EK, Kristan M, Irish SR, Walker T: **Evidence for natural hybridization and novel Wolbachia strain superinfections in the *Anopheles gambiae* complex from Guinea.** *Roy Soc Open Sci* 2021, **8**(4).
52. Ramirez-Puebla ST, Ormeno-Orrillo E, de Leon AVP, Lozano L, Sanchez-Flores A, Rosenblueth M, Martinez-Romero E: **Genomes of *Candidatus Wolbachia bourtzisii* wDacA and *Candidatus Wolbachia pipientis* wDacB from the Cochineal Insect *Dactylopius coccus* (Hemiptera: Dactylopiidae).** *G3-Genes Genom Genet* 2016, **6**(10):3343-3349.
53. Shropshire JD, Bordenstein SR: **Two-By-One model of cytoplasmic incompatibility: Synthetic recapitulation by transgenic expression of cifA and cifB in *Drosophila*.** *Plos Genet* 2019, **15**(6).
54. Hui Y, Nanxian Z, Xiaocheng J, Yizhu C, Wenquan Z: **Reproductive characters of *Ficus hirta* Vahl (Moraceae) and its symbiotic fig wasps.** *Chinese Bulletin of Botany* 2004, **21**(6):682-688.
55. Weiblen GD: **How to be a fig wasp.** *Annu Rev Entomol* 2002, **47**:299-330.
56. Yu H, Zhao NX, Chen YZ, Herre EA: **Male and female reproductive success in the dioecious fig, *Ficus hirta* Vahl. in Guangdong Province, China: Implications for the relative stability of dioecy and monoecy.** *Symbiosis* 2008, **45**(1-3):121-127.
57. Greeff JM, van Noort S, Rasplus JY, Kjellberg F: **Dispersal and fighting in male pollinating fig wasps.** *Cr Biol* 2003, **326**(1):121-130.
58. Yu H, Nong XJ, Fan SL, Bhanumas C, Deng XX, Wang R, Chen XY, Compton SG: **Olfactory and gustatory receptor genes in fig wasps: Evolutionary insights from comparative studies.** *Gene* 2023, **850**.
59. Dadd R: **Nutrition: organisms.** *Comprehensive insect physiology, biochemistry and pharmacology* 1985, **4**:313-390.
60. Wood HG, Barden RE: **Biotin enzymes.** *Annual review of biochemistry* 1977, **46**(1):385-413.
61. Knowles JR: **The mechanism of biotin-dependent enzymes.** *Annual review of biochemistry* 1989, **58**(1):195-221.
62. Hu R, Sun P, Yu H, Cheng Y, Wang R, Chen X, Kjellberg F: **Similitudes and differences between two closely related *Ficus* species in the synthesis by the ostiole of odors attracting their host-specific pollinators: A transcriptomic based investigation.** *Acta oecologica* 2020, **105**:103554.
63. Yu H, Nason JD, Zhang L, Zheng LN, Wu W, Ge XJ: **De novo transcriptome sequencing in *Ficus hirta* Vahl. (Moraceae) to investigate gene regulation involved in the biosynthesis of pollinator attracting volatiles.** *Tree Genet Genomes* 2015, **11**(5).
64. Bublitz DC, Chadwick GL, Magyar JS, Sandoz KM, Brooks DM, Mesnage S, Ladinsky MS, Garber AI, Bjorkman PJ, Orphan VJ *et al*: **Peptidoglycan Production by an Insect-Bacterial Mosaic.** *Cell* 2019, **179**(3):703+.
65. Zhu DT, Rao Q, Zou C, Ban FX, Zhao JJ, Liu SS: **Genomic and transcriptomic analyses reveal metabolic complementarity between whiteflies and their symbionts.** *Insect Sci* 2022, **29**(2):539-549.
66. Ju JF, Bing XL, Zhao DS, Guo Y, Xi ZY, Hoffmann AA, Zhang KJ, Huang HJ, Gong JT, Zhang X *et al*: **Wolbachia supplement biotin and riboflavin to enhance reproduction in planthoppers.** *Isme J* 2020, **14**(3):676-687.
67. Chaisson MJ, Tesler G: **Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory.** *Bmc Bioinformatics* 2012, **13**.
68. Kolmogorov M, Yuan J, Lin Y, Pevzner PA: **Assembly of long, error-prone reads using repeat graphs.** *Nat Biotechnol* 2019, **37**(5):540+.
69. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng QD, Wortman J, Young SK *et al*: **Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement.** *Plos One* 2014, **9**(11).
70. Shen W, Le S, Li Y, Hu FQ: **SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation.** *Plos One* 2016, **11**(10).
71. Varani AM, Siguier P, Goubeyre E, Charneau V, Chandler M: **ISSaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes.** *Genome Biol* 2011, **12**(3).
72. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J: **eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale.** *Mol Biol Evol* 2021, **38**(12):5825-5829.
73. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M: **KAAS: an automatic genome annotation and pathway reconstruction server.** *Nucleic Acids Research* 2007, **35**:W182-W185.
74. Rozewicki J, Li SL, Amada KM, Standley DM, Katoh K: **MAFFT-DASH: integrated protein sequence and structural alignment.** *Nucleic Acids Research* 2019, **47**(W1):W5-W10.

75. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS: **ModelFinder: fast model selection for accurate phylogenetic estimates**. *Nat Methods* 2017, **14**(6):587-+.
76. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ: **IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies**. *Mol Biol Evol* 2015, **32**(1):268-274.
77. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S: **High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries**. *Nat Commun* 2018, **9**.

Tables

Table 1 Different reads screened by BLASR and assembled with different parameters of Flye

BLASR parameter*	Total length of reads (bp)	Flye parameter*	Genome size after assembly (bp)	Number of Fragments
Identity $\geq$ 60%, maxscore $\leq$ -250	1,024,923,876	minimum overlap between reads $\geq$ 1000 bp	53,393,107	1561
		minimum overlap between reads $\geq$ 3000 bp	53,568,170	1501
		minimum overlap between reads $\geq$ 5000 bp	53,930,492	1476
Identity $\geq$ 60%, maxscore $\leq$ -300	639,517,259	minimum overlap between reads $\geq$ 1000 bp	1,734,532	70
		minimum overlap between reads $\geq$ 3000 bp	2,258,584	55
		minimum overlap between reads $\geq$ 5000 bp	2,352,926	50

Note * The asterisk indicates that other command parameters of BLASR are the same. The command for BLASR is blasr input.fasta db.fasta -out outfile -unaligned unalignedfilename -nproc 8 -bestn 1 -nCandidates 10 -minPctIdentity 60 -minReadLength 500 -maxScore -250 -maxLCPLength 15.

Table 2 Genome Statistics of *Wolbachia* in *Valisia javana*

Parameters	<i>w.Jav</i>
Host	<i>Valisia javana</i>
Infection type	Multiple infection, two different strains
Total size (bp)	2,352,827
Number of contigs	50
GC content (%)	35.0
Number of CDS	2825
Number of RNAs	62

Table 3 MLST sequence of *Wolbachia* in *Valisia javana*

query sequence	Identity	length	evaluate	Housekeeping gene name (Allele)
fig 953.597.peg.158	98.113	371	0	<i>gatB</i> (78)
fig 953.597.peg.671	99.317	439	0	<i>fbpA</i> (62)
fig 953.597.peg.810	99.507	406	0	<i>coxA</i> (1) 1 difference found. 51T → 576G
fig 953.597.peg.874	99.754	406	0	<i>coxA</i> (1)
fig 953.597.peg.1067	96.305	406	0	<i>coxA</i> (37)
fig 953.597.peg.1260	97.732	441	0	<i>hcpA</i> (60)
fig 953.597.peg.1667	99.772	439	0	<i>ftsZ</i> (3)
fig 953.597.peg.2329	100	371	0	<i>gatB</i> (1)
fig 953.597.peg.2642	99.772	439	0	<i>ftsZ</i> (3)
fig 953.597.peg.2669	99.317	439	0	<i>fbpA</i> (62)

Note: The genes with two sequences accession numbers are indicated by boldface type. The query sequences could refer to Additional file 9.

Table 4 Different MLST allelic profiles combinations of *wJav1* and *wJav2*, and unique MLST allelic profiles of five closed related species.

	Strains	<i>gatB</i>	<i>coxA</i>	<i>hcpA</i>	<i>ftsZ</i>	<i>fbpA</i>	ST	host species	supergroup
group 1	<i>wJav1</i>	1	1	60	3	62	-	<i>Valisia javana</i>	-
	<i>wJav2</i>	78	37	60	3	62	-	<i>Valisia javana</i>	-
group 2	<i>wJav1</i>	1	37	60	3	62	-	<i>Valisia javana</i>	-
	<i>wJav2</i>	78	1	60	3	62	-	<i>Valisia javana</i>	-
	A_M11He	54	52	62	3	62	83	unknown	A
	Sole_A_17-02	78	37	41	3	62	122	<i>Solenopsis</i>	A
	Mchi_A	78	37	90	3	62	123	unknown	A
	A_NY_Calyp150744c	78	37	62	3	62	132	unknown	A
	A_Pan_Droso150749b	93	37	93	3	62	133	unknown	A

Table 5 The closest Average Nucleotide Identity (ANI) of *wJav1* and *wJav2* with fifteen *Wolbachia* strains

Assembly accession of reference sequence	Assembly name	ANI (%)	host species	Strain	supergroup
GCA_018690095.1	ASM1869009v1	98.9087	<i>Drosophila simulans</i>	wAu	A
GCA_000742435.1	ASM74243v1	98.9024	<i>Drosophila recens</i>	wRec	A
GCF_014129525.1	ASM1412952v1	98.8729	<i>Drosophila tropicalis</i>	wTro	A
GCF_016584375.1	ASM1658437v1	98.7803	<i>Drosophila melanogaster</i>	wMelOctoless	A
GCF_016584325.1	ASM1658432v1	98.751	<i>Drosophila melanogaster</i>	wMelpop	A
GCF_016584355.1	ASM1658435v1	98.751	<i>Drosophila melanogaster</i>	wMelpop2	A
GCF_016584405.1	ASM1658440v1	98.7484	<i>Drosophila melanogaster</i>	wMelCS_b	A
GCF_017896365.1	ASM1789636v1	98.7267	<i>Aedes aegypti</i>	GV_2018_1	A
GCF_007972595.1	ASM797259v1	98.7229	<i>Drosophila melanogaster</i>	wMel_I23	A
GCF_016584425.1	ASM1658442v1	98.7121	<i>Drosophila melanogaster</i>	wMel	A
GCF_014129655.1	ASM1412965v1	98.701	<i>Drosophila</i>	wAra	A
GCA_007971685.1	ASM797168v1	98.685	<i>Drosophila melanogaster</i>	wMel_N25	A
GCF_017896305.1	ASM1789630v1	98.6849	<i>Aedes aegypti</i>	GV_2018_5	A
GCA_021497225.1	ASM2149722v1	98.6847	<i>Aedes aegypti</i>	wMel_GV_13	A
GCF_000008025.1	ASM802v1	98.6845	<i>Drosophila melanogaster</i>	wMel	A

Table 6 Genes related to biotin synthesis and metabolism annotated in the *wJav* genome

Gene ID	COG category	KEGG ko	PFAMs	Description	Preferred_name
fig 953.597.peg.606	C	ko:K00627	2-oxoacid_dh, Biotin_lipoyl, E3_binding	The pyruvate dehydrogenase complex catalyzes the overall conversion of pyruvate to acetyl-CoA and CO(2)	<i>pdhC</i>
fig 953.597.peg.285	C	ko:K00658	2-oxoacid_dh, Biotin_lipoyl, E3_binding	The 2-oxoglutarate dehydrogenase complex catalyzes the overall conversion of 2-oxoglutarate to succinyl-CoA and CO(2)	<i>sucB</i>
fig 953.597.peg.373	M	ko:K02022,ko:K12542	Biotin_lipoyl_2, HlyD,HlyD_3	Secretion protein	<i>aprE</i>
fig 953.597.peg.1008	M	ko:K02022,ko:K12542	Biotin_lipoyl_2, HlyD,HlyD_3	Secretion protein	<i>aprE</i>
fig 953.597.peg.1587	H	ko:K03524	BPL_C, BPL_LplA_LipB, BPL_LplA_LipB_2	Biotin/lipoate A/B protein ligase family	<i>birA</i>
fig 953.597.peg.1588	H	ko:K03524	BPL_C, BPL_LplA_LipB, BPL_LplA_LipB_2	Biotin/lipoate A/B protein ligase family	<i>birA</i>
fig 953.597.peg.1204	I	ko:K01965	Biotin_carb_C, Biotin_carb_N, Biotin_lipoyl,CPSase_L_D2	Propionyl-CoA carboxylase alpha	<i>pccA</i>
fig 953.597.peg.1314	C	ko:K00658	2-oxoacid_dh, Biotin_lipoyl, E3_binding	The 2-oxoglutarate dehydrogenase complex catalyzes the overall conversion of 2-oxoglutarate to succinyl-CoA and CO(2)	<i>sucB</i>
fig 953.597.peg.2169	M	ko:K02022,ko:K12542	Biotin_lipoyl_2, HlyD,HlyD_3	Secretion protein	<i>aprE</i>
fig 953.597.peg.2777	H	ko:K03524	BPL_C, BPL_LplA_LipB, BPL_LplA_LipB_2	Biotin/lipoate A/B protein ligase family	<i>birA</i>
fig 953.597.peg.2495	H	ko:K03524	BPL_C,BPL_LplA_LipB,BPL_LplA_LipB_2	Biotin/lipoate A/B protein ligase family	<i>birA</i>
fig 953.597.peg.2535	M	ko:K18990	Biotin_lipoyl_2, HlyD_D23	Belongs to the membrane fusion protein (MFP) (TC 8.A.1) family	-

Figures

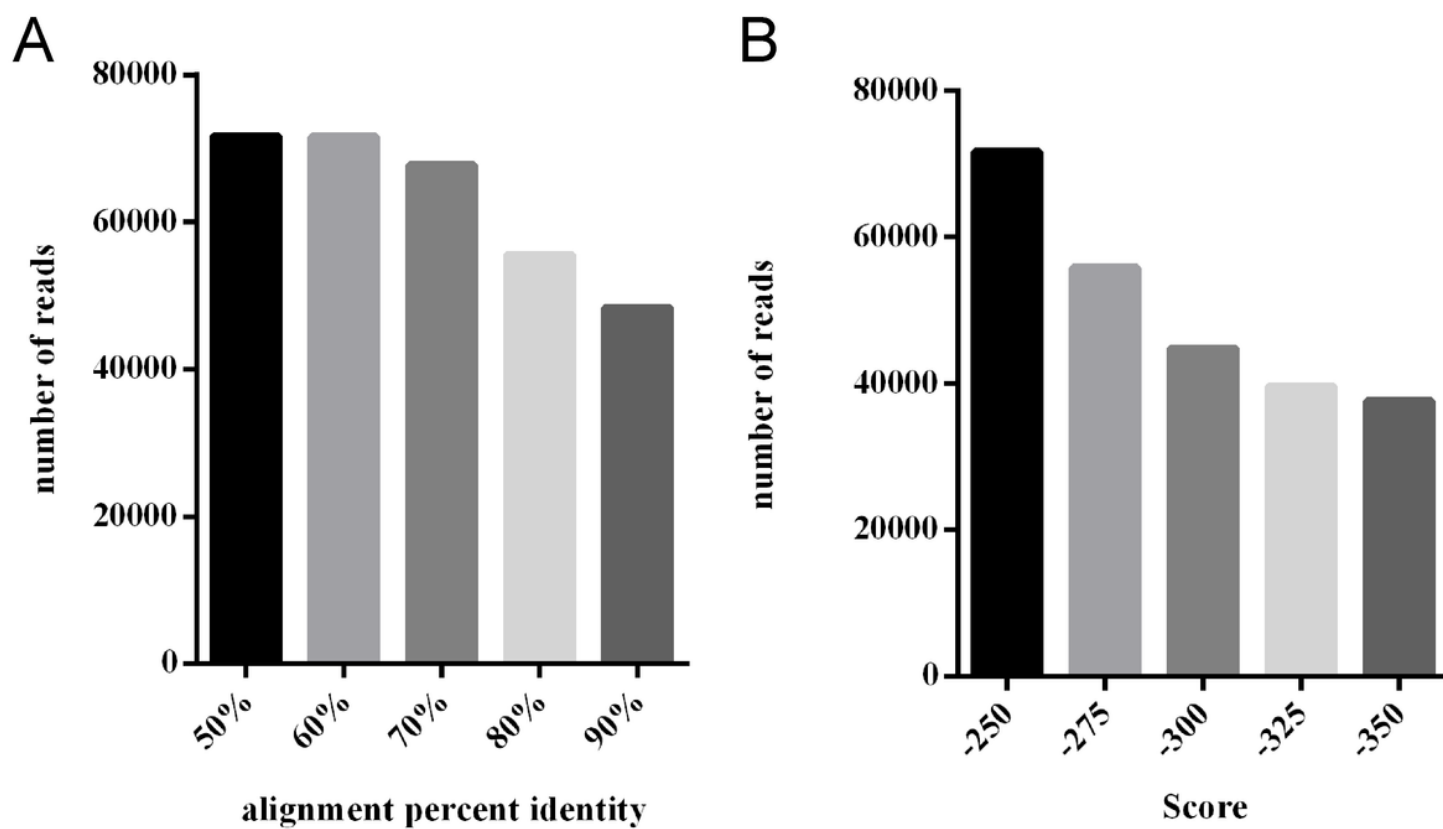


Figure 1

The number of *Wolbachia* reads aligned by BLASR in the host genome Pacbio sequencing. (A) The number of *Wolbachia* genome reads at different alignment percent identity. (B). The number of *Wolbachia* genome reads at different scores.

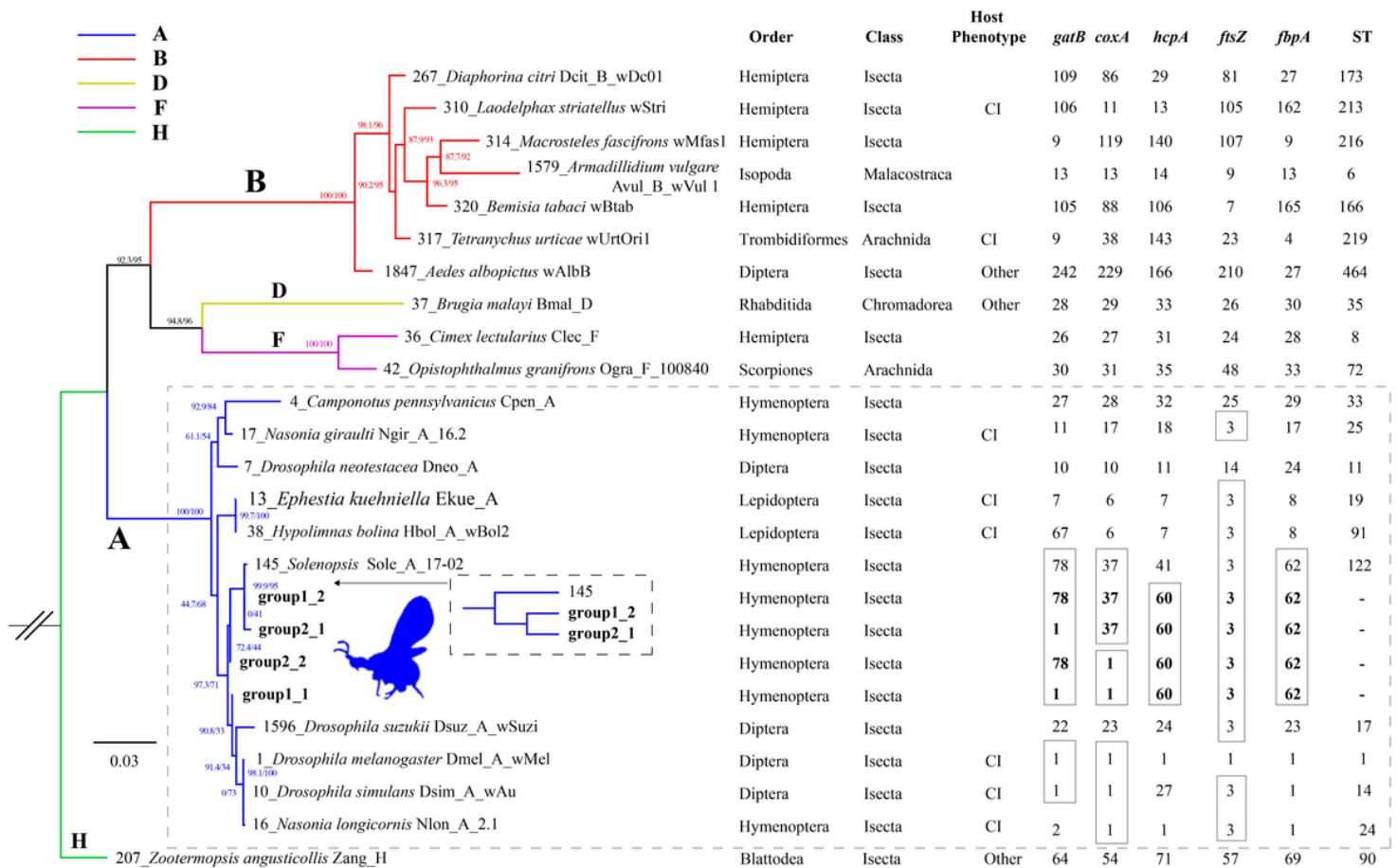


Figure 2

Phylogenetic analysis of 21 *Wolbachia* MLST data belong to different supergroups based on maximum-likelihood. SH-aLRT and UFBoot support values are given on nodes. Corresponding sample ids, hosts, strains and colors represent different supergroups of *Wolbachia* strains could be referred to Table S2. MLST groups could refer to Additional file 9.

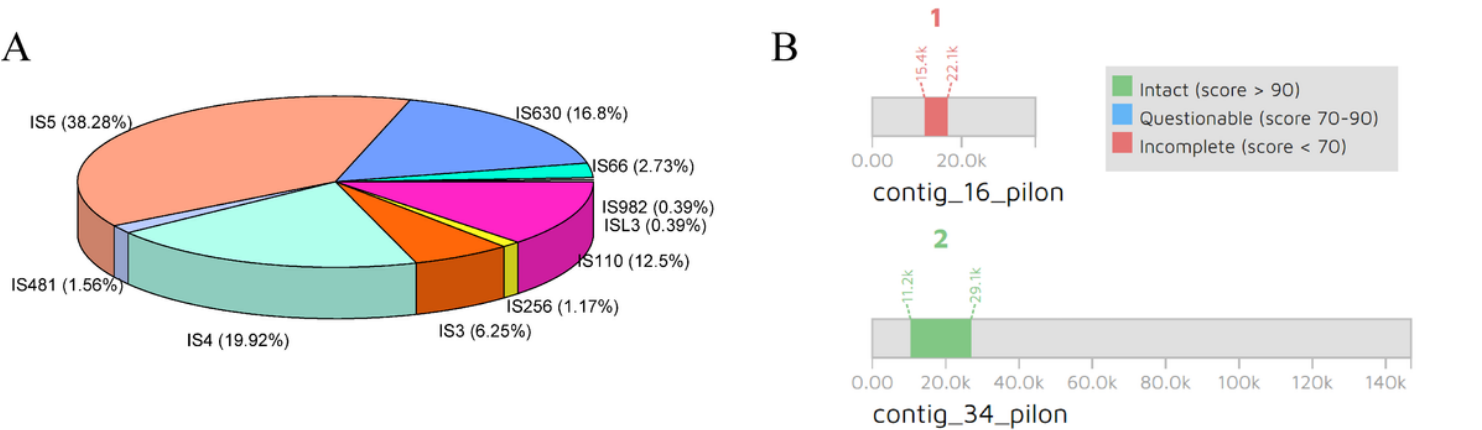


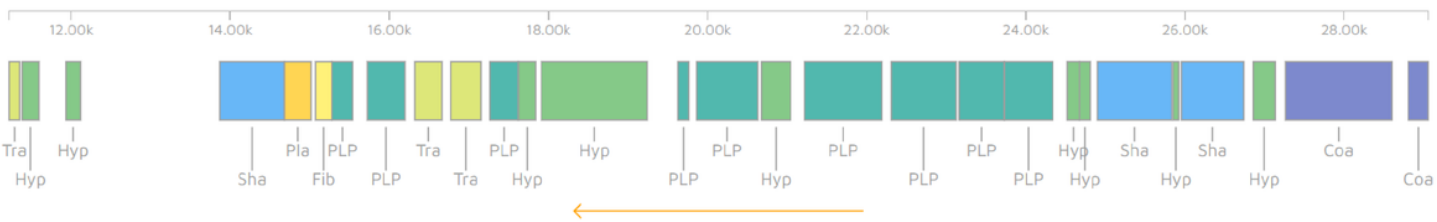
Figure 3

Mobile genetic elements situation summary in the mixed genome of *wJav1* and *wJav2*. (A) Proportion of ten main types of insertion sequence families. (B) Regions of two prophages referred to as WOjav1-2 in the genome (see Figure 4 and Table S5 for details). The integrity of the prophage regions was marked by different color.

WOJav1



WOJav2



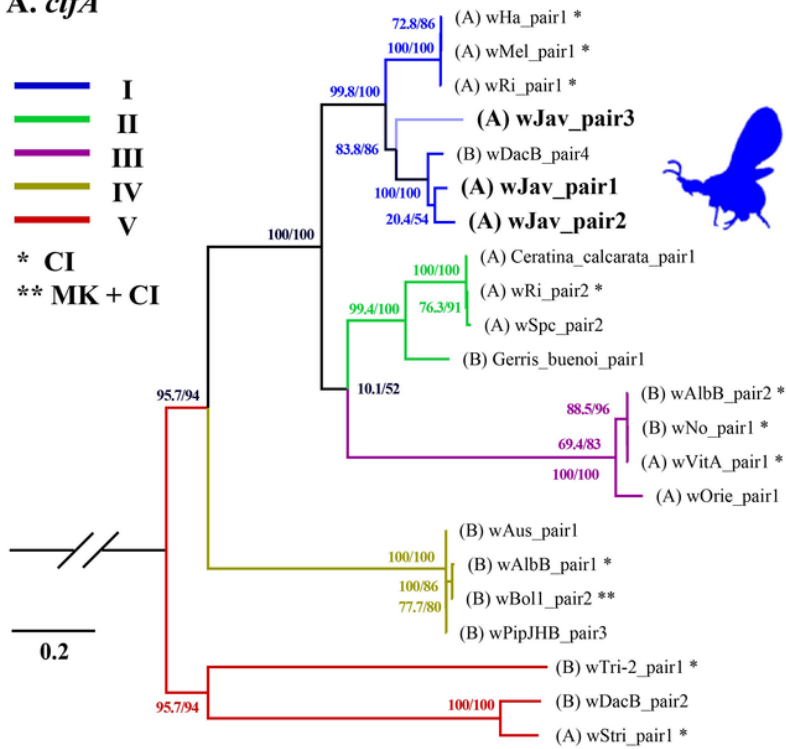
Lysis	Protease	Attachment Site	Hypothetical Protein	Fiber Protein
Terminase	Coat Protein	Integrase	Other	Plate Protein
Portal Protein	Tail Shaft	Phage-like Protein	Transposase	tRNA

Figure 4

Prophage regions detected in the mixed genome of *wJav1* and *wJav2* using the PHASTER server. The number of genes discovered in each prophage region is shown in the lower panel with different color. See Table S5 for details

A. *cifA*

- I
- II
- III
- IV
- V
- * CI
- ** MK + CI



B. *cifB*

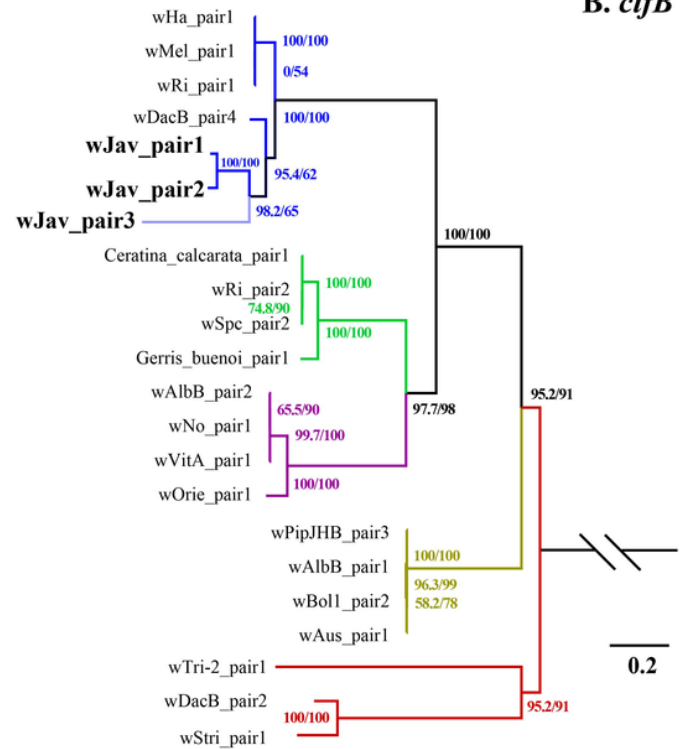


Figure 5

ML trees of (A) *cifA* and (B) *cifB* genes pairs in *wJav1-2* genomes and 19 *Wolbachia* strains. SH-aLRT and UFBoot support values are given on nodes. Supergroup of the *Wolbachia* strains are given in parentheses. * refers to CI host phenotype. ** refers to CI and MK. Corresponding colors that represent different *cif* types could be referred to Table S4. Corresponding strains name could be referred to Table S6 and Table S7.

BIOTIN METABOLISM

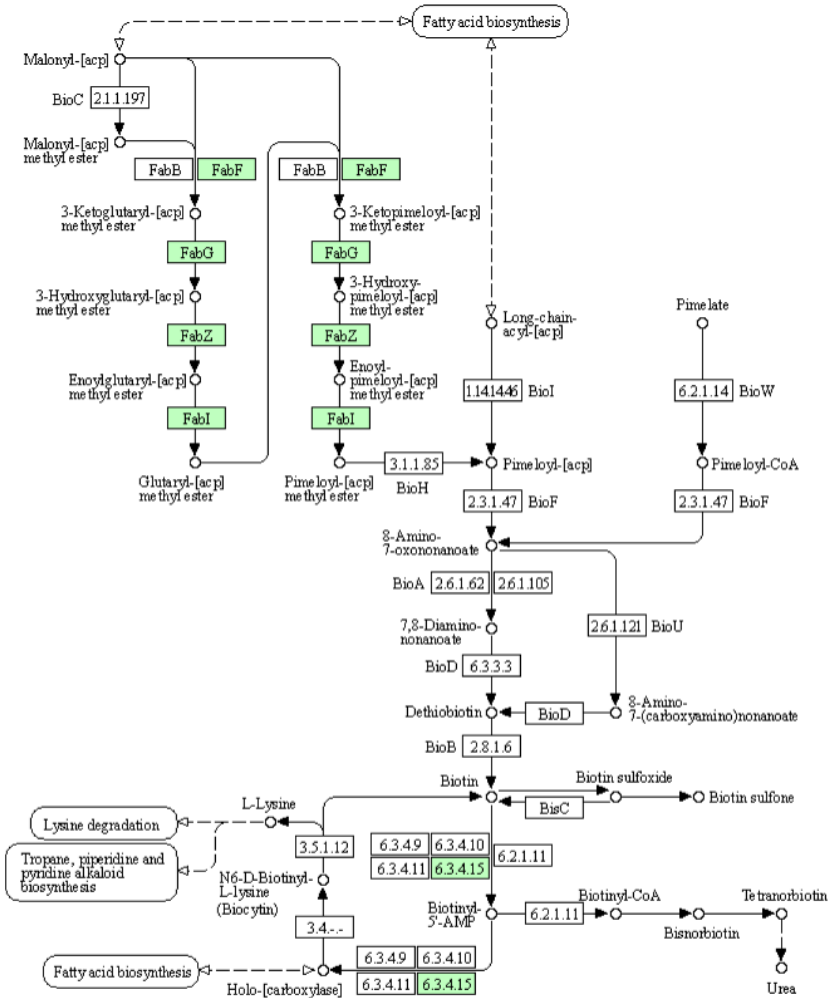


Figure 6

A schematic illustration of genes related to biotin synthesis and metabolism in *wJav1-2*. Numbers in boxes indicate Enzyme Commission numbers. White and green boxes indicate absence and presence of the corresponding enzymes, respectively, based on GhostKoala annotation. See Table 6 for details.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfiles19.xls