Miao et al. *Zool. Res.* 2023, 44(2): 273−275
https://doi.org/10.24272/j.issn.2095-8137.2022.315

ZR | Zoological Research

Letter to the editor Open Access

Complicated expansion trajectories of insertion sequences and potential association with horizontal transfer of *Wolbachia* DNA

DEAR EDITOR,

Insertion sequences (ISs) are the simplest structural transposable elements (TEs) in prokaryotes, consisting only of a transposase coding sequence and its bilateral short terminal inverted repeats. Due to their gradually streamlined genomic construction, TEs rarely exist in the genomes of obligate endosymbionts. However, TE content, especially ISs, is abundant in the genome of *Wolbachia* bacteria, obligate endosymbionts widespread in arthropods and nematodes. Although IS indels are reported to affect genome structure and gene function in *Wolbachia*, the distribution patterns, sources, and transfer trajectories of ISs remain poorly understood. Furthermore, whether IS transposition is associated with dynamic horizontal transfer of *Wolbachia* DNA is still unclear. Based on distribution patterns in supergroup A *Wolbachia* strains, ISs accounted for 11% of the genome of the *Wolbachia* strain *w*Wpum, one of the highest IS genome coverages reported for *Wolbachia* to date. Three types of ISs showed rapid expansion in *w*Wpum, possibly due to horizontal transfer from other *Wolbachia* strain supergroups or more distant prokaryotes. We also found the first evidence that ISs can carry flanking *Wolbachia* sequences for transposition, resulting in the horizontal transfer of *Wolbachia* DNA into the eukaryotic genome, thus implying a potential association between ISs and horizontal gene transfer from endosymbionts to eukaryotes.

TEs are heritable DNA sequences that move freely from one site to another, even across species, and exist widely in both eukaryotic and prokaryotic genomes (Bourque et al., 2018). Prokaryotes usually have smaller genomes, less TE content, and fewer TE types than eukaryotes. ISs are the simplest structural TEs in prokaryotes, with a length of no more than 2 500 bp, usually consisting of a transposase coding sequence and its bilateral short terminal inverted repeats (IR). Certain IS types can form two short direct repeats (DR) in their terminals when inserted into target sites. More than 4 000 types of ISs in 29 families have been identified in prokaryotes based on transposase similarity, conservation of catalytic sites, and IR similarity (Siguier et al., 2006).

It is commonly accepted that obligate endosymbionts are restricted to the cell, obtaining nutrients from the host rather than self-synthesis, resulting in streamlined genomes with few TEs and coding regions only for essential proteins (Wernegreen, 2002). However, *Wolbachia* bacteria, maternally inherited obligate endosymbionts with complicated phylogenetic relationships (including 18 lineages (supergroups A–R) in the sole species) and widespread distribution in arthropods and nematodes, contain abundant TEs, especially ISs. For instance, IS content in the *Wolbachia* strain *w*Ri, which infects *Drosophila simulans*, exceeds 10%, the highest IS genome coverage reported in prokaryotes to our knowledge (Cerveau et al., 2011).

Wolbachia genomes are often affected by ISs. For example, based on genomic comparison of the *Wolbachia* strains *w*Ri and *w*Mel, approximately half of the gene-order breakpoints are flanked by ISs (Klasson et al., 2009). Genomic rearrangements caused by ISs are also reported in the *w*Suzi strain (Kaur et al., 2017) and inversion of *Wolbachia* genome fragments mediated by ectopic recombination between IRs of ISs appears to be relatively common (Ling & Cordaux, 2010), providing a reasonable explanation for the complex genomic rearrangements in *Wolbachia*. *Wolbachia* strains also exhibit active horizontal gene transfer (HGT), with more than 70% of eukaryotic hosts estimated to have HGT from *Wolbachia* (Hotopp, 2011). Although TE transposition can cause HGT (Sieber et al., 2017), in-depth studies on IS transfer and expansion and the potential association between IS transposition and HGT in *Wolbachia* are lacking.

Here, we selected seven circularized genome assemblies of diverse *Wolbachia* strains in supergroup A from GenBank (Supplementary Table S1) to identify IS distribution patterns at the genome-wide level. A series of subsequent analyses were performed (see detailed methods in Supplementary Materials and Supplementary Table S2–S8). In total, 905 IS copies involving 10 IS families were identified from the seven *Wolbachia* genomes, of which 672 copies were assigned to known IS groups (Supplementary Table S9). IS content varied markedly among the strains. Notably, IS content accounted for 11% of the genome of the *w*Wpum strain infecting the fig wasp *Wiebesia pumilae* (Hymenoptera), one of the highest IS genome coverages reported in prokaryotes.

Three IS types were significantly expanded in *w*Wpum

Received: 27 October 2022; Accepted: 03 January 2023; Online: 07 January 2023

Foundation items: This study was supported by the National Natural Science Foundation of China (31830084, 32070466) and Fundamental Research Funds for the Central Universities, Nankai University (96172158, 96173250, 91822294)

compared to the other six strains, including ISWpi4 (*n*=22, *P*=0.008, one copy in each other strain) in the IS481 family, ISOt6 (*n*=15, absent in other strains) in the IS5 family, and an unnamed IS type (*n*=50, also in *w*CauA (*n*=53) but absent in others) in the IS5 family. We named the unnamed IS type ISWpi19 following naming conventions of ISs in *Wolbachia*. The three IS types accounted for more than half of the total IS length (75 078/139 363 bp=53.87%) in *w*Wpum, thus explaining the record-setting IS content. We also constructed gene trees of IS types to investigate their evolutional relationships. Each IS member was named according to the positional order of its locus on the genome assembly (from 5' to 3'). All ISWpi4 members in the seven *Wolbachia* strains fell into two clades (Supplementary Figure S1A). Clade β consisted of 21 ISWpi4 copies (except for ISWpi4-*w*Wpum11) from *w*Wpum, with an average identity of 99.98% (Supplementary Figure S2A), suggesting that they arose via a recent expansion event. Similarly, the average identity of the 15 ISOt6 copies was 99.89% (Supplementary Figure S2B). ISWpi19-*w*Wpum and ISWpi19-*w*CauA were clustered separately (Supplementary Figure S1B), and copies within the same clade showed little divergence, with average identities of 99.87% (ISWpi19-*w*Wpum) and 99.98% (ISWpi19-*w*CauA) (Supplementary Figure S2C), suggesting that the expansion of ISWpi19 in *w*Wpum and *w*CauA occurred via two independent events. Based on the divergence time tree of the supergroup A *Wolbachia* strains (Supplementary Figure S1C), ISWpi19 expanded in *w*CauA but was absent in the other strain (*w*Ha) of the same clade, indicating that ISWpi19-*w*CauA expansion occurred later than the divergence between *w*CauA and *w*Ha, within about 10 000 years, and ISWpi19-*w*Wpum may have expanded more recently due to its higher average identity than ISWpi19-*w*CauA.

Given that the three IS types were rarely distributed in other closely related supergroup A strains, we searched GenBank for homologous sequences of each IS type and constructed divergence time trees to trace their expansion trajectories. ISWpi19 homologous sequences were discovered in Cyanobacteria, Euryarchaeotes, Bacteroidetes, and *Wolbachia* strains (Supplementary Figure S3A). In the clade of *Wolbachia* origin, ISWpi19-*w*Wpum and ISWpi19-*w*CauA copies identified from the supergroup A strains were closer to homologous sequences of supergroup B strains than other supergroup A strains, indicating that they may be derived directly from strains in supergroup B. ISWpi4 homologous sequences were identified in 27 *Wolbachia* strains and *Aphantopus hyperantus* (Lepidoptera) (Supplementary Figure S3B). Although most homologous sequences were distributed in supergroup A, the most divergent sequence was from the *w*VulC strain in supergroup B, and an independent clade from the non-supergroup A strain *w*CfeJ occurred in the supergroup A clades, implying HGT of ISWpi4 among diverse *Wolbachia* supergroups. ISOt6 homologous sequences were distributed across multiple classes of prokaryotes but in only two *Wolbachia* strains, *w*Nik and *w*Wpum. They clustered with the *Parachlamydia acanthamoebae* sequence into one clade (clade γ) (Supplementary Figure S3C). According to multilocus sequence typing (MLST), *w*Nik was not phylogenetically closely related to *w*Wpum (Supplementary Figure S4), indicating that ISOt6 may have been horizontally transferred to *Wolbachia* from other distant prokaryotes.

Interestingly, the only non-*Wolbachia* ISWpi4 homolog (ISWpi4 region) (Supplementary Figure S3B) was located on chromosome 7 of *A. hyperantus* (GenBank accession No.: LR761654.1), flanked by two highly homologous sequences to *Wolbachia* (upstream and downstream regions of ISWpi4). We named the entire sequence containing these three regions as sequence X (Figure 1A). The AT base content of sequence X was 67.37%, significantly higher than that of its 20 kb upstream (63.13%, *P*<0.0001) and downstream (64.61%, *P*=0.0448) flanking regions (Figure 1B), and even higher than that of the *A. hyperantus* genome (62%), implying exogeneity of sequence X. We also constructed divergence time trees for the upstream and downstream regions in sequence X (Figure 1C) and found that the upstream region occurred much later in *A. hyperantus* than in other eukaryotes but closer to the *Wolbachia* strains. Furthermore, the downstream region occurred later in *A. hyperantus* than in many *Wolbachia* strains, with no eukaryotic homolog discovered. Thus, these results suggest that sequence X in the eukaryote *A. hyperantus* was not of simple eukaryotic origin but was derived directly from *Wolbachia* HGT.

The same TCAGA sequence (5 bp) was identified in both terminals of sequence X and the 3'-terminal of the ISWpi4 region, but not in its 5'-terminal (CTAGG) (Figure 1A). We speculated that HGT of sequence X from *Wolbachia* to *A. hyperantus* was related to ISWpi4 transposition. Thus, we compared the insertion site characteristics of ISWpi4-*w*Wpum and ISWpi4-*w*OneA1 copies closely related to the ISWpi4 region in sequence X. As expected, the TCAGA and CTAGG sequences were present in the terminals of the ISWpi4-*w*Wpum and ISWpi4-*w*OneA1 copies, respectively. Each IS copy insertion site shared at least 60% similarity (Supplementary Figure S5), implying that TCAGA and CTAGG are hot spots for ISWpi4 insertion. We also identified five identical ISOt6 copies in strain *w*Nik with highly homologous flanking sequences (*Wol*-homolog1 and *Wol*-homolog2, respectively) (Supplementary Figure S6). *Wol*-homolog1 was present in four ISOt6-*w*Nik copies but absent in ISOt6-*w*Nik2. Similarly, *Wol*-homolog2 was present in four ISOt6-*w*Nik copies, but absent in ISOt6-*w*Nik3, with different lengths. No *Wol*-homolog1 or *Wol*-homolog2 homologs were identified elsewhere in the *w*Nik genome. Interestingly, two groups of IRs were discovered in the full-length sequence of "*Wol*-homolog1+ISOt6+*Wol*-homolog2", including two internal IRs of ISOt6 (17 bp) (IR-up/down) and three IRs (44 bp) in the flanking regions (IR-A/B/C). If ISOt6 recognizes the random two of five IRs as cleavage sites for transposition, the products would match the patterns of the five sequences in Supplementary Figure S6. These results suggest that ISs carrying flanking *Wolbachia* DNA for transposition is a common occurrence.

In conclusion, this study revealed the recent distantly related origin expansion of ISs in *w*Wpum, as well as the first evidence of a possible association between IS transposition and HGT from *Wolbachia* to eukaryotes. Nevertheless, given the isolated evidence, how ISs transfer from other distant prokaryotes to *Wolbachia* remains unclear. Therefore, more abundant genomic data are needed to solve these questions and deepen our understanding of the genomic evolution of symbiotic systems.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Y.H.M., D.W.H., and J.H.X. designed the study. Y.H.M. analyzed the data.

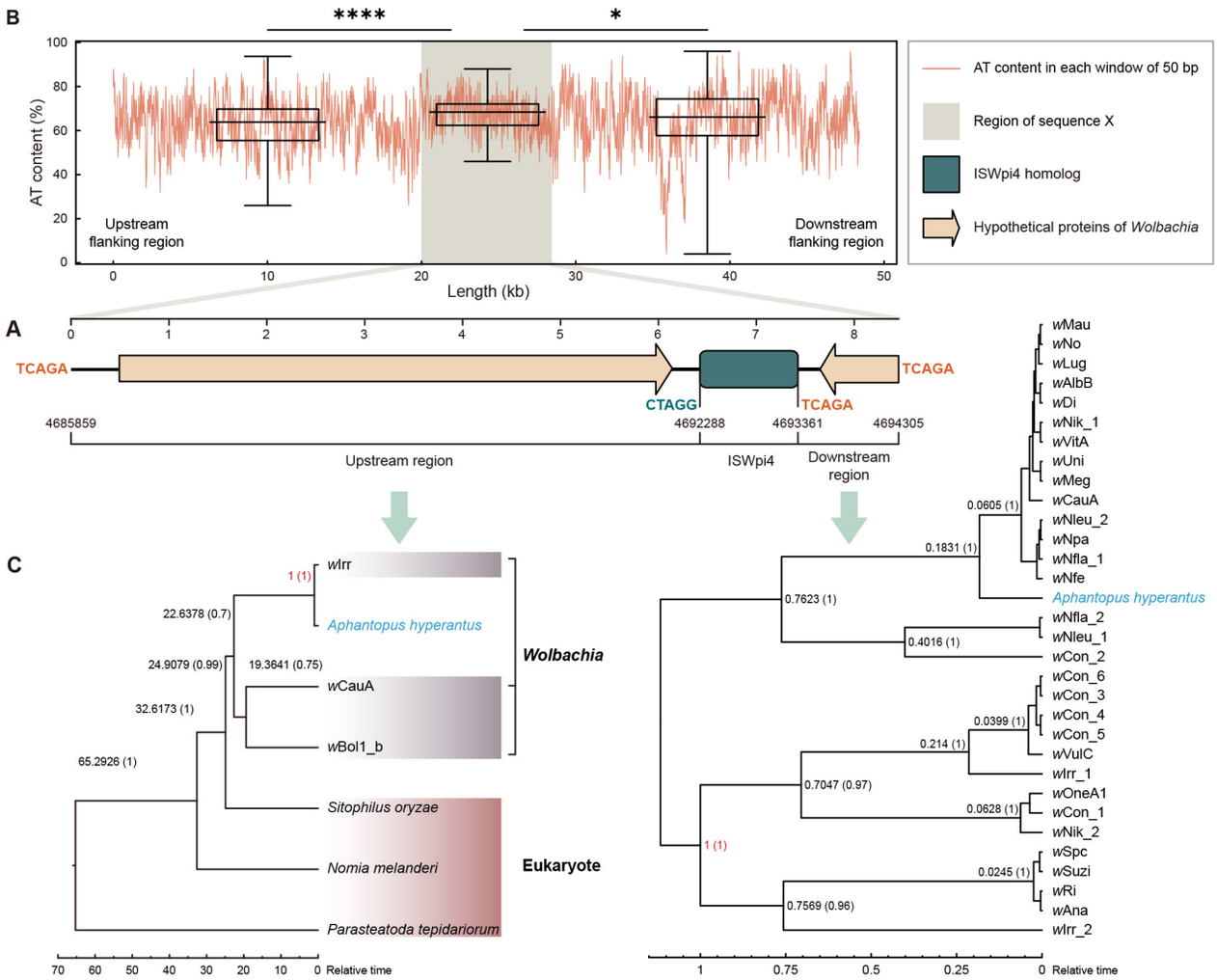


Figure 1 *Wolbachia* origin of sequence X in *A. hyperantus*

A: Schematic structure of sequence X was divided into three parts artificially, including ISWpi4 region and its upstream and downstream regions. Numbers on bottom axis show location of the three parts on chromosome 7. Orange and green nucleotides are bases in terminals of sequence X and ISWpi4 region. B: AT content in sequence X and its flanking regions was calculated for each 50 bp window, ranging 20 kb upstream and 20 kb downstream of sequence X. Mann-Whitney *U*-test, *: $P<0.05$; ****: $P<0.0001$. C: Relative divergence trees for homologs of upstream (left) and downstream regions (right) of ISWpi4 in sequence X. Leaves of sequence X are marked as "*Aphantopus hyperantus*" in blue. Average divergence time of each main node in the trees is labeled (with posterior probability in brackets), and nodes set as relative age of 1 are in red.

Y.H.M. and J.H.X. wrote the manuscript. All authors read and approved the final version of the manuscript.

Yun-Heng Miao[1], Da-Wei Huang[1,*], Jin-Hua Xiao[1,*]

[1] *College of Life Sciences*, *Nankai University*, *Tianjin* 300071, *China*

*Corresponding authors, E-mail: huangdw@ioz.ac.cn; xiaojh@nankai.edu.cn

REFERENCES

Bourque G, Burns KH, Gehring M, et al. 2018. Ten things you should know about transposable elements. *Genome Biology*, **19**(1): 199.

Cerveau N, Leclercq S, Leroy E, et al. 2011. Short- and long-term evolutionary dynamics of bacterial insertion sequences: insights from Wolbachia endosymbionts. *Genome Biology and Evolution*, **3**: 1175−1186.

Hotopp JCD. 2011. Horizontal gene transfer between bacteria and animals. *Trends in Genetics*, **27**(4): 157−163.

Kaur R, Siozios S, Miller WJ, et al. 2017. Insertion sequence polymorphism and genomic rearrangements uncover hidden *Wolbachia* diversity in *Drosophila suzukii* and *D. subpulchrella*. *Scientific Reports*, **7**(1): 14815.

Klasson L, Westberg J, Sapountzis P, et al. 2009. The mosaic genome structure of the *Wolbachia wRi* strain infecting *Drosophila simulans*. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(14): 5725–5730.

Ling A, Cordaux R. 2010. Insertion sequence inversions mediated by ectopic recombination between terminal inverted repeats. *PLoS One*, **5**(12): e15654.

Sieber KB, Bromley RE, Hotopp JCD. 2017. Lateral gene transfer between prokaryotes and eukaryotes. *Experimental Cell Research*, **358**(2): 421−426.

Siguier P, Perochon J, Lestrade L, et al. 2006. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Research*, **34**(S1): D32−D36.

Wernegreen JJ. 2002. Genome evolution in bacterial endosymbionts of insects. *Nature Reviews Genetics*, **3**(11): 850−861.

ZR Zoological Research

Supplementary Materials

Complicated expansion trajectories of insertion sequences and potential association with horizontal transfer of *Wolbachia* DNA

Yun-Heng Miao[1], Da-Wei Huang[1,*], Jin-Hua Xiao[1,*]

[1] College of Life Sciences, Nankai University, Tianjin 300071, China

[*]Corresponding authors, E-mail: huangdw@ioz.ac.cn; xiaojh@nankai.edu.cn

Supplementary Materials and Methods

Genomic data acquisition of *Wolbachia*

To explore the insertion sequence (IS) distribution patterns at the genome-wide level, we selected circularized genome assemblies of seven *Wolbachia* strains in supergroup A, including the *w*Wpum strain published by our laboratory and six *Wolbachia* strains (*w*Ana, *w*Au, *w*CauA, *w*Ha, *w*Mel, and *w*Ri) published in GenBank (Supplementary Table S1). The phylogenetic relationship among the seven *Wolbachia* strains was relatively close, which was conducive to subsequent analysis and avoided highly complex differences among genomes due to large genetic distances.

Supplementary Table S1 Seven *Wolbachia* assemblies analyzed in the study

Strain	Accession Number	Host	Reference
*w*Wpum	CP054557.1	*Wiebesia pumilae*	(Miao et al., 2020)
*w*Mel	CP046925.1	*Drosophila melanogaster*	(Wu et al., 2004)
*w*Au	NZ_LK055284.1	*Drosophila simulans*	(Sutton et al., 2014)
*w*Ana	CP042904.1	*Drosophila ananassae*	
*w*Ri	CP001391.1	*Drosophila simulans*	(Lisa et al., 2009)
*w*CauA	CP041215.1	*Carposina sasakii*	
*w*Ha	NC_021089.1	*Drosophila simulans*	(Ellegaard et al., 2013)

Identification and annotation of ISs

We identified ISs from the *Wolbachia* genomes using three strategies according to previous proposed methods (Cerveau et al., 2011).

Strategy 1: Prodigal v2.6.3 (Hyatt et al., 2010) was used to predict the coding sequences of proteins in the genomes. BLASTP was then performed to compare the proteins to the NR database in GenBank (*e*-value≤1e-5) and proteins were obtained using annotation keywords “transposase” and “transposon”. The locus of each protein sequence was recorded.

Strategy 2: Similarity searches were performed against the IS reference database ISfinder (Siguier et al., 2006) by BLASTN (*e*-value≤1e-5) to predict ISs in the genomes. The locus of each IS copy was recorded.

Strategy 3: *De novo* repeat detection was performed for each genome using RepeatScout v1.0.5 (Price et al., 2005) with l-mer size=15 bp. The RepeatScout results were used as queries for BLASTN against the NR database to identify non-IS repeats (e.g., phages, group II introns, duplicated genes). All repeats with identities over 60% to non-IS repeats were discarded. Finally, the streamlined results were BLASTN (*e*-value≤1e-5) searched against the *Wolbachia* genomes to obtain sequences and corresponding loci of each IS copy in the genome.

For each *Wolbachia* genome, we annotated all sequences detected by the above strategies using TBlastX (*e*-value≤1e-5) against ISfinder to obtain the IS-family and IS-group to which they belong. Next, we manually eliminated potentially redundant or overlapping ISs in two steps. Firstly, we considered sequences assigned to the same IS-family and covering the same locus range as one IS copy and recorded the locus of the sequence with the broadest coverage as the IS copy locus. Secondly, when overlapping sequences were assigned to the same IS-family and IS-group and each was part of the same IS reference, they were considered as one IS copy.

When overlapping sequences were not assigned to the same IS-family, we extracted their full-range sequence as a query for BLASTN against ISfinder to confirm whether the query matched a specific IS reference. If it did, those sequences were considered as one IS copy. Otherwise, the sequences were retained separately.

Construction of gene trees for IS types with significant expansion

For each significantly expanded IS type, we used MAFFT v7.313 (Katoh & Standley, 2013) to perform multiple sequence alignment against all IS copies. We then used Gblocks v0.91b (Talavera & Castresana, 2007) to select conserved blocks in the alignments. MrBayes v3.2.7a (Altekar et al., 2004) was used to construct the gene tree with the optimal nucleotide substitution model predicted using ModelFinder (Kalyaanamoorthy et al., 2017).

Estimation of divergence time of 27 *Wolbachia* strains in supergroup A

In addition to the 27 *Wolbachia* strains in supergroup A, we also selected one strain in supergroup B (*w*No) as the outgroup (Supplementary Table S2). The coding proteins of each strain were predicted using Prodigal v2.6.3, with OrthoMCL (Li et al., 2003) (default parameters) then used to search for orthologous proteins among the 28 strains. All single-copy orthologous proteins in each strain were concatenated in a specific order into a super-protein sequence. The 28 super-protein sequences were aligned, and the conserved blocks were selected as described above. The phylogenetic tree was constructed using MrBayes with the optimal amino acid substitution model predicted using ProtTest v3.4.2 (Darriba et al., 2011).

Supplementary Table S2 *Wolbachia* genomes (*n*=28) used for chronogram construction

Supergroup	Strain	Accession Number	Host
A	*w*CsolI	CP054598	*Ceratosolen Solmsi*
A	*w*Kgib	JABXYD000000000	*Kradibia gibbosae*
A	*w*Wpum	CP054557.1	*Wiebesia pumilae*
A	*w*San	NZ_VCEH00000000.1	*Drosophila santomea*
A	*w*Tei	NZ_VCEG00000000.1	*Drosophila teissieri*
A	*w*Yak	NZ_VCEF00000000.1	*Drosophila yakuba*
A	*w*Mel	NC_002978.6	*Drosophila melanogaster*
A	*w*MelPop	NZ_AQQE00000000.1	*Drosophila melanogaster*
A	*w*Inc_Cu	CP011148.1	*Drosophila incompta*
A	*w*Inc_SM	CP011149.1	*Drosophila incompta*
A	*w*Au	NZ_LK055284.1	*Drosophila simulans*
A	*w*Rec	NZ_JQAM00000000.1	*Drosophila recens*
A	*w*Ana	NZ_CP042904.1	*Drosophila ananassae*
A	*w*Ri	CP001391.1	*Drosophila simulans*
A	*w*Spc	NZ_NTHL00000000.1	*Drosophila subpulchrella*
A	*w*Suzi	NZ_CAOU00000000.2	*Drosophila suzukii*
A	*w*OneA1	NZ_QESS00000000.1	*Nasonia oneida*
A	*w*CauA	CP041215.1	*Carposina sasakii*
A	*w*Ha	NC_021089.1	*Drosophila simulans*
A	*w*Uni	NZ_MUJL01000001.1	*Muscidifurax uniraptor*

A	*w*VitA	NZ_MUJM01000001.1	*Nasonia vitripennis*
A	*w*Gmo	NZ_AWUH00000000.1	*Glossina morsitans*
A	*w*Nfla	NZ_LYUW00000000.1	*Nomada flava*
A	*w*Nleu	NZ_LYUV00000000.1	*Nomada leucophthalma*
A	*w*Npa	NZ_LYUX00000000.1	*Nomada panzeri*
A	*w*DacA	NZ_LSYX00000000.1	*Dactylopius coccus*
A	*w*Nfe	NZ_LYUY00000000.1	*Nomada ferruginata*
B	*w*No	NC_021084.1	*Drosophila simulans*

To estimate divergence times of the 28 strains, BEAST v1.10.4 (Suchard et al., 2018) was used to construct time trees, with the MCMC chain length set to 100 000 000. Six trees were generated using different models (Bayesian Skyline, Constant Size, and Exponential Growth models with uncorrelated relaxed clock type and strict clock type, respectively) and the optimal tree was selected based on the PS/SS MLE test (Supplementary Table S3). To calibrate the divergence times of *Wolbachia* strains, we selected two published divergence time points, i.e., divergence time between *w*Mel and *w*MelPop of 1 153 years ago (95% confidence interval 445–2 041 years ago) (Richardson et al., 2012) and divergence time between *w*Suzi and *w*Spc of about 1 000–10 000 years ago (Conner et al., 2017).

Supplementary Table S3 PS/SS MLE test results for divergence time trees of six different models.

Clock Type	Model	PS	SS
Uncorrelated relaxed clock	Bayesian Skyline	-512405.17	-512398.14
	Constant Size	-512411.69	-512409.52
	Exponential Growth	-512412.73	-512409.76
Strict clock	Bayesian Skyline	-513371.20	-513364.55
	Constant Size	-513316.30	-513304.49
	Exponential Growth	-513315.97	-513304.20

Table note: Highest PS/SS score was selected as the optimal model.

Source trace of IS types with significant expansion in *w*Wpum and sequence X in *Aphantopus hyperantus*

We used the three IS types showing significant expansion (ISWpi19, ISOt6, and ISWpi4) in the *Wolbachia* strain *w*Wpum and two regions (upstream and downstream region of ISWpi4) of sequence X in *Aphantopus hyperantus* as queries to perform a BLASTN search against the NR database in GenBank. The search results were used as the sequence sets to construct divergence time trees, regardless of whether the set members were intact ISs.

(a) For ISWpi19, we selected search results with *e*-value≤1e-10 and identity coverage ≥20% to ISWpi19, ISMac15, and ISCaa9 (ISMac15 and ISCaa9 are two IS references similar to ISWpi19 in ISfinder) as the sequence set. To simplify calculation, we only chose the sequence with the highest BLAST score for each species or *Wolbachia* strain (Supplementary Table S4).

Supplementary Table S4 Genomes involved in chronogram of ISWpi19 homologs

	Strain	Accession Number
Cyanobacteria	***Calothrix* sp. NIES-3974**	**AP018254.1**
Cyanobacteria	***Nostoc* sp. C057**	**CP040289.1**
Cyanobacteria	**Nostoc flagelliforme CCNUN1**	**CP024785.1**
Euryarchaeotes	**Methanosarcinales archaeon ANME-1 ERB6**	**MT631450.1**
Euryarchaeotes	**Methanosarcina horonobensis HB-1**	**CP009516.1**
Euryarchaeotes	**Methanosarcina vacuolata Z-761**	**CP009520.1**
Euryarchaeotes	**Methanosarcina sp. Kolksee**	**CP009524.1**
Euryarchaeotes	**Methanosarcina barkeri str. Wiesmoor**	**CP009526.1**
Euryarchaeotes	**Methanosarcina mazei C16**	**CP009514.1**
Euryarchaeotes	**Methanosarcina sp. WWM596**	**CP009503.1**
Euryarchaeotes	**Methanosarcina siciliae C2J**	**CP009508.1**
Euryarchaeotes	**Methanosarcina acetivorans C2A**	**AE010299.1**
Bacteroidetes	**Candidatus Amoebophilus asiaticus 5a2**	**CP001102.1**
Wolbachia	***w*Ana**	**NZ_CP042904.1**
Wolbachia	***w*Ri**	**CP001391.1**
Wolbachia	***w*Ha**	**NC_021089.1**
Wolbachia	***w*Mel**	**NC_002978.6**
Wolbachia	***w*Inc_Cu**	**CP011148.1**
Wolbachia	***w*Inc_SM**	**CP011149.1**
Wolbachia	***w*Au**	**NZ_LK055284.1**
Wolbachia	***w*Irr**	**CP037426.1**
Wolbachia	***w*Stv**	**CP050531.1**
Wolbachia	***w*Wpum**	**CP054557.1**
Wolbachia	***w*CauA**	**CP041215.1**
Wolbachia	**dawsonii**	**CP051608.1**
Wolbachia	***w*PipPel**	**NC_010981.1**
Wolbachia	***w*AlbB**	**CP031221.1**
Wolbachia	**China 1**	**CP016430.1**
Wolbachia	**wCfeJ**	**CP051157.1**

(b) For ISOt6, we selected search results with *e*-value≤1e-10 and identity coverage ≥70% to ISOt6. As in (a), we only chose one sequence with the highest BLAST score for each *Wolbachia* strain (Supplementary Table S5).

Supplementary Table S5 Genomes involved in chronogram of ISOt6 homologs

Strain	Accession Number
Allofrancisella frigidaquae strain SYSU 10HL1970	CP038017.1
Allofrancisella guangzhouensis strain 08HL01032	CP010427.1
Allofrancisella inopinata strain SYSU YG23	CP038241.1

Candidatus Amoebophilus asiaticus 5a2	CP001102.1
Candidatus Fukatsuia symbiotica strain 5D	CP021659.1
Candidatus Fukatsuia symbiotica strain 5D plasmid p5D_Fsymbiotica-1	CP021660.1
Candidatus Fukatsuia symbiotica strain 5D plasmid p5D_Fsymbiotica-3	CP021662.1
Candidatus Hamiltonella defensa strain ZA17	CP017613.1
Candidatus Paracaedibacter acanthamoebae isolate PRA3	CP008941.1
Candidatus Regiella insecticola clone CUGI_APP_BA-03-N07	AC192956.2
Legionella anisa isolate UMCG_3A	CP029563.1
Legionella anisa isolate UMCG_3A plasmid p3A1	CP029564.1
Legionella pneumophila subsp. pascullei strain D-7119	CP014257.1
Legionella pneumophila subsp. pascullei strain D-7158	CP014256.1
Legionella pneumophila subsp. pascullei strain F-4185	CP014255.2
Legionella pneumophila subsp. pascullei strain NCTC12272	LS483412.1
Legionella pneumophila subsp. pascullei strain NCTC12273	LR134380.1
Legionella pneumophila subsp. pascullei strain U8W (D-7160)	CP021262.1
Legionella sainthelensi strain NCTC11988	LR134388.1
Legionella spiritensis strain NCTC11990	LT906457.1
Neochlamydia sp. S13 DNA	AP017977.1
Orientia tsutsugamushi Boryong	AM494475.1
Orientia tsutsugamushi isolate Karp	LS398548.1
Orientia tsutsugamushi isolate Kato	LS398550.1
Orientia tsutsugamushi isolate TA686	LS398549.1
Orientia tsutsugamushi isolate UT176	LS398547.1
Orientia tsutsugamushi isolate UT76	LS398552.1
Orientia tsutsugamushi str. Gilliam	LS398551.1
Orientia tsutsugamushi str. Ikeda DNA	AP008981.1
Orientia tsutsugamushi strain Wuj/2014	CP044031.1
Parachlamydia acanthamoebae UV-7	FR872580.1
Regiella insecticola BAC CUGI_APP_BA-001B13	AC202204.3
Regiella insecticola BAC CUGI_APP_BA-001H17	AC202203.2
Regiella insecticola BAC CUGI_APP_BA-001K06	AC202195.2
Regiella insecticola BAC CUGI_APP_BA-001P06	AC203073.1
Regiella insecticola BAC CUGI_APP_BA-012D01	AC202208.4
Regiella insecticola BAC CUGI_APP_BA-012L18	AC202197.2
Regiella insecticola BAC CUGI_APP_BA-01-D22	AC192954.2
Regiella insecticola BAC CUGI_APP_BA-03-B02	AC192955.1
Spartinivicinus ruber strain S2-4-1H	CP048878.1
*и*Nik	CP050530.1
*и*Wpum	CP054557.1

(c) For ISWpi4, we selected all homologs with identity coverage ≥90% to ISwpi4 (Supplementary Table S6).

Supplementary Table S6 Genomes involved in chronogram of ISWpi4 homologs

	Strain / Spieces	Accession Number
Wolbachia	*w*Kgib	JABXYD000000000
Wolbachia	*w*Wpum	CP054557.1
Wolbachia	*w*San	NZ_VCEH00000000.1
Wolbachia	*w*Tei	NZ_VCEG00000000.1
Wolbachia	*w*Yak	NZ_VCEF00000000.1
Wolbachia	*w*Mel	NC_002978.6
Wolbachia	*w*MelPop	NZ_AQQE00000000.1
Wolbachia	*w*Inc_Cu	CP011148.1
Wolbachia	*w*Inc_SM	CP011149.1
Wolbachia	*w*Au	NZ_LK055284.1
Wolbachia	*w*Rec	NZ_JQAM00000000.1
Wolbachia	*w*Ana	NZ_CP042904.1
Wolbachia	*w*Ri	CP001391.1
Wolbachia	*w*Spc	NZ_NTHL00000000.1
Wolbachia	*w*Suzi	NZ_CAOU00000000.2
Wolbachia	*w*OneA1	NZ_QESS00000000.1
Wolbachia	*w*CauA	CP041215.1
Wolbachia	*w*Ha	NC_021089.1
Wolbachia	*w*Uni	NZ_MUJL01000001.1
Wolbachia	*w*VitA	NZ_MUJM01000001.1
Wolbachia	*w*Nfla	NZ_LYUW00000000.1
Wolbachia	*w*Nleu	NZ_LYUV00000000.1
Wolbachia	*w*Npa	NZ_LYUX00000000.1
Wolbachia	*w*Nfe	NZ_LYUY00000000.1
Wolbachia	*w*VulC	NZ_ALWU01000001.1
Wolbachia	*w*CfeJ	CP051157.1
Wolbachia	*w*Irr	CP037426.1
Eukaryote	*Aphantopus hyperantus*	LR761654.1

(d) For the upstream region of ISWpi4 in sequence X, we selected sequences with identity coverage ≥70% and *e*-value≤1e-10 (Supplementary Table S7).

Supplementary Table S7 Genomes involved in chronogram of homologs of upstream region in sequence X

	Strain / Spieces	Accession Number
Wolbachia	*w*Irr	CP037426.1

Wolbachia	*w*CauA	CP041215.1
Wolbachia	*w*Bol1-b	NZ_CAOH00000000.1
Coleoptera	*Sitophilus oryzae*	XM_030890056.1
Hymenoptera	*Nomia melanderi*	XM_031992809.1
Araneae	*Parasteatoda tepidariorum*	XM_030890056.1

(e) For the downstream region of ISWpi4 in sequence X, we selected results with identity coverage ≥90% (Supplementary Table S8).

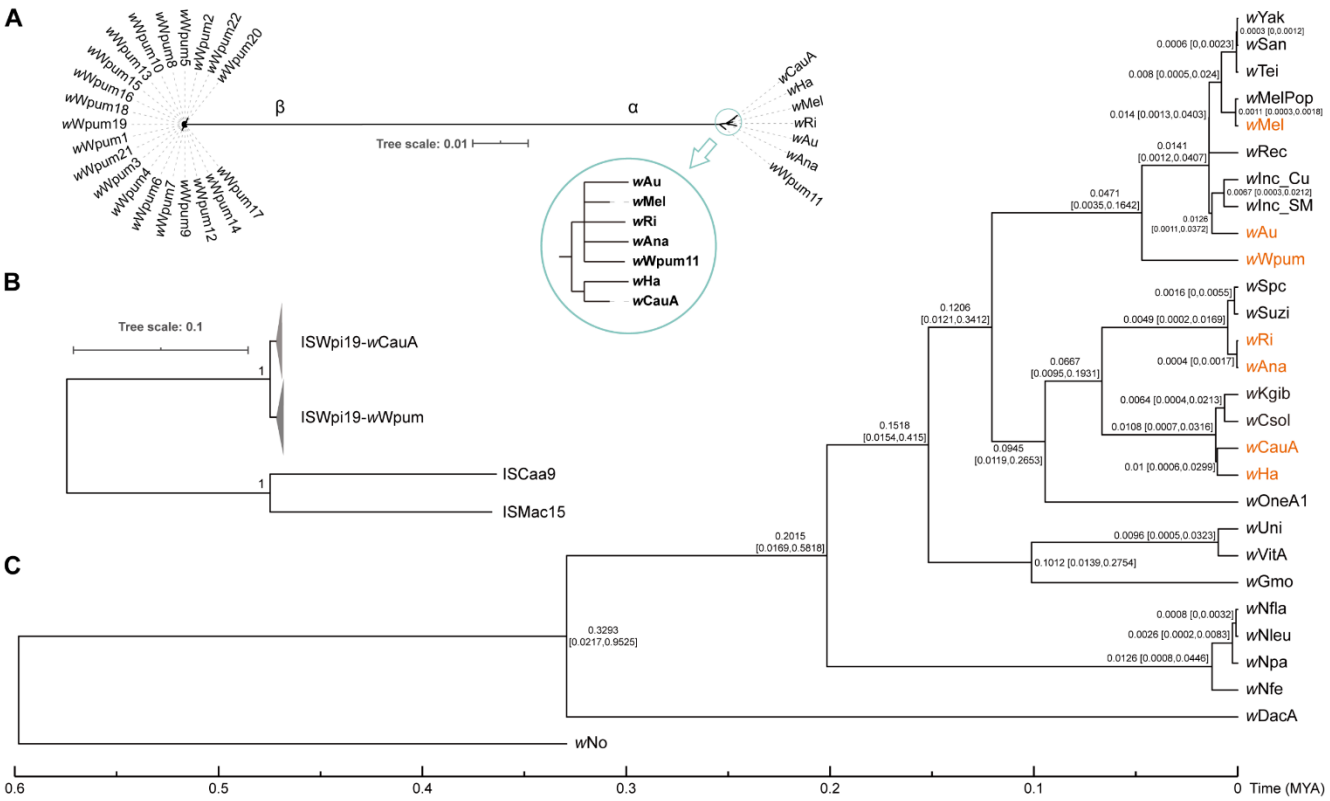
Supplementary Table S8 Genomes involved in chronogram of homologs of downstream region in sequence X

	Strain / Spieces	Accession Number
Wolbachia	*w*Mau	CP034335.1
Wolbachia	*w*No	NC_021084.1
Wolbachia	*w*AlbB	CP031221.1
Wolbachia	*w*Lug	NZ_MUIY00000000.1
Wolbachia	*w*Di	NZ_AMZJ01000001.1
Wolbachia	*w*Nik	CP050530.1
Wolbachia	*w*Uni	NZ_MUJL01000001.1
Wolbachia	*w*VitA	NZ_MUJM01000001.1
Wolbachia	*w*Meg	CP021120.1
Wolbachia	*w*CauA	CP041215.1
Wolbachia	*w*Nfla	NZ_LYUW00000000.1
Wolbachia	*w*Nleu	NZ_LYUV00000000.1
Wolbachia	*w*Npa	NZ_LYUX00000000.1
Wolbachia	*w*Nfe	NZ_LYUY00000000.1
Wolbachia	*w*Con	NZ_QPIP00000000.1
Wolbachia	*w*OneA1	NZ_QESS00000000.1
Wolbachia	*w*VulC	NZ_ALWU01000001.1
Wolbachia	*w*Irr	CP037426.1
Wolbachia	*w*Ri	CP001391.1
Wolbachia	*w*Spc	NZ_NTHL00000000.1
Wolbachia	*w*Suzi	NZ_CAOU00000000.2
Wolbachia	*w*Ana	CP042904.1
Eukaryote	*Aphantopus hyperantus*	LR761654.1

We used BEAST to construct time trees using Yule models with uncorrelated relaxed clock type and strict clock type, respectively, and the MCMC chain length was set to 100 000 000. The tree with higher posterior probability was considered optimal. Given the limited fossil evidence and reports for calibration of divergence time, we set the relative time of a certain node to 1 to construct arbitrary scaling time trees following previous study (Conner et al., 2017).

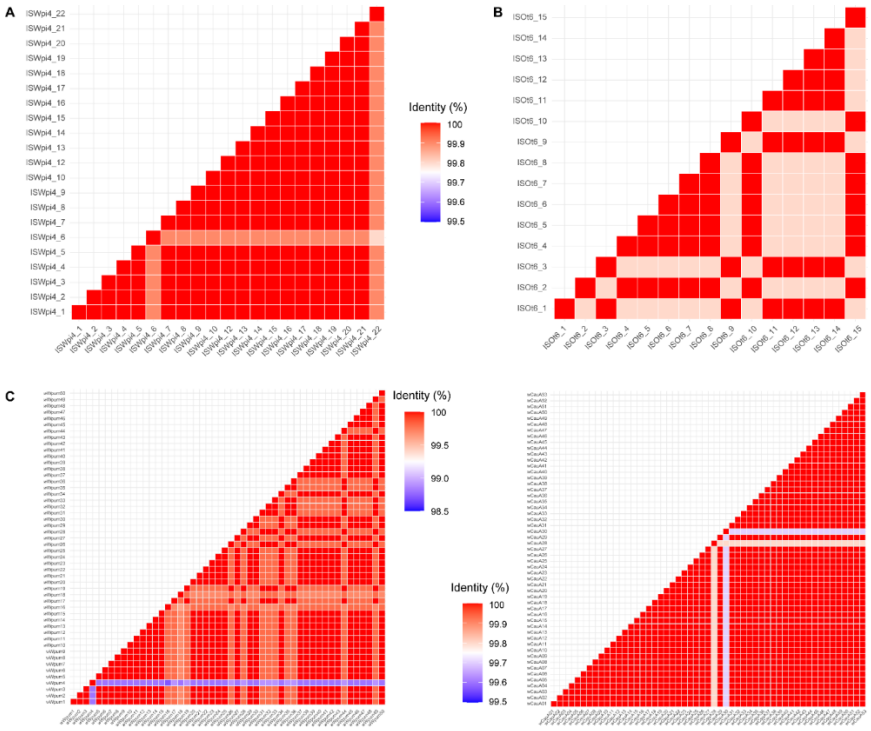
Supplementary Table S9 Statistics of ISs in seven *Wolbachia* strains. Group column "UC" means "Unclassified", indicating that ISs in that row were not classified to IS-group level. "Proportion" indicates percentage of IS length accounting for genome length of *Wolbachia*.

Family	Group	*w*Wpum	*w*Mel	*w*Au	*w*Ana	*w*Ri	*w*CauA	*w*Ha
IS110		**9**	**8**	**9**	**29**	**29**	**12**	**11**
	IS1111	1	1	1	1	1	3	1
	UC	8	7	8	28	28	9	10
IS256		**3**	**3**	**3**	**1**	**1**	**2**	**2**
	UC	3	3	3	1	1	2	2
IS3		**22**	**15**	**15**	**13**	**14**	**21**	**18**
	IS3	20	15	14	12	13	13	15
	IS150	0	0	0	0	0	5	0
	UC	2	0	1	1	1	3	3
IS4		**17**	**15**	**15**	**17**	**11**	**16**	**20**
	IS231	2	3	3	3	3	4	5
	IS4	7	7	7	13	7	7	12
	IS50	7	5	5	1	1	5	3
IS481		**22**	**1**	**1**	**11**	**10**	**4**	**4**
	UC	22	1	1	11	10	4	4
IS5		**102**	**27**	**29**	**24**	**29**	**68**	**13**
	IS1031	69	17	16	21	26	59	4
	IS903	18	10	13	3	3	9	9
	ISL2	15	0	0	0	0	0	0
IS630		**5**	**5**	**5**	**8**	**10**	**8**	**7**
	UC	5	5	5	8	10	8	7
IS66		**7**	**11**	**7**	**37**	**35**	**10**	**11**
	ISBst12	7	11	7	37	35	10	11
IS982		**3**	**1**	**1**	**1**	**1**	**1**	**1**
	UC	3	1	1	1	1	1	1
ISNCY		**5**	**5**	**5**	**3**	**3**	**9**	**4**
	ISPlu15	5	5	5	3	3	9	4
Total		193	91	90	144	143	151	91
IS Length (bp)		139,363	65,571	62,578	136,450	145,166	105,277	62,396
Genome Length (bp)		1,267,465	1,267,782	1,268,461	1,401,460	1,445,873	1,449,344	1,295,804
Proportion		11.00%	5.17%	4.93%	9.74%	10.04%	7.26%	4.82%



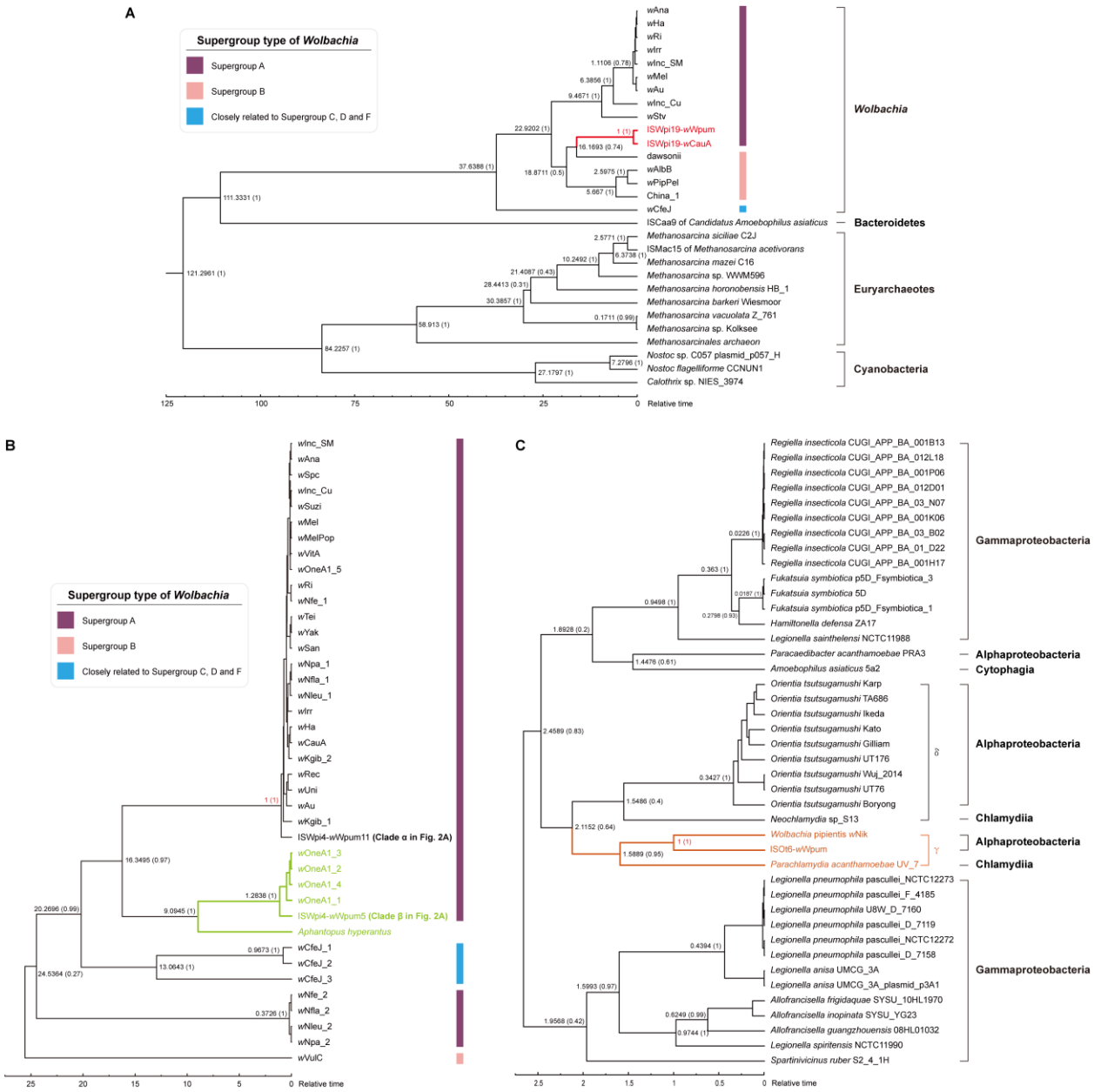
Supplementary Figure S1 Expansion patterns of three IS types and chronogram for *Wolbachia* strains in supergroup A

A: Gene tree of ISWpi4 members in seven *Wolbachia* strains constructed with nucleotide substitution model HKY+F and no outgroup setting. Phylogenetic relationship of members in α lineage is magnified in green circle. Posterior probability of each main node was 1. **B:** Gene tree of ISWpi19 members in *w*Wpum and *w*CauA constructed with nucleotide substitution model HKY+F, with two reference ISs most similar to ISWpi19 (ISCaa9 and ISMac15) in ISfinder set as outgroups. **C:** Divergence time tree of 27 *Wolbachia* strains in supergroup A using time calibration of two nodes (*w*Mel: *w*MelPop and *w*Spc: *w*Suzi), with *w*No in supergroup B as the outgroup, and amino acid substitution model of CpRev+I+G. Estimated divergence time of each node is shown with 95% confidence intervals with posterior probability of 1.



Supplementary Figure S2 Identity between each two IS elements

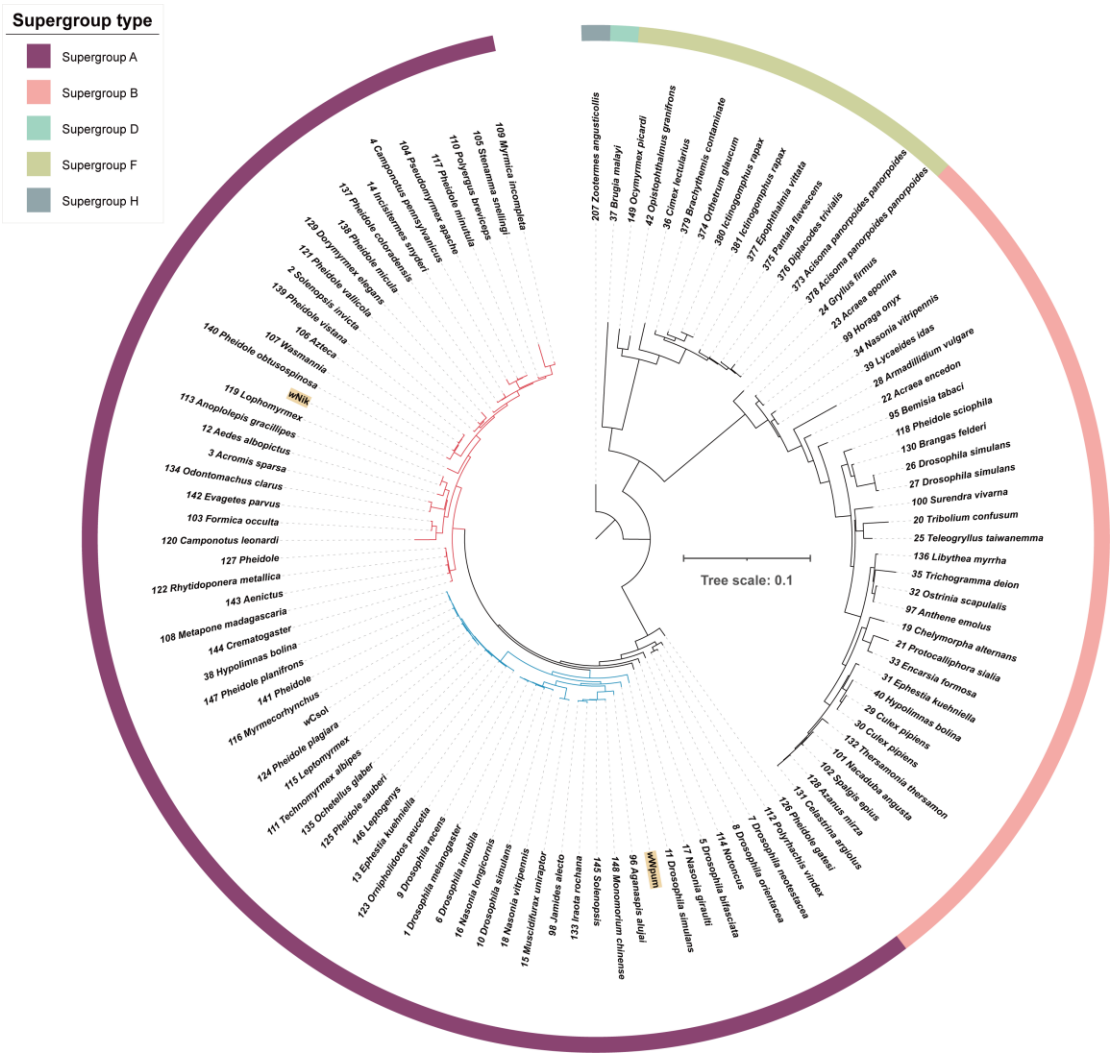
A: ISWpi4-*w*Wpum, **B:** ISOt6-*w*Wpum, and **C:** ISWpi19-*w*Wpum (left) and ISWpi19-*w*CauA (right).



Supplementary Figure S3 Chronograms of three IS types expanded in *w*Wpum genome

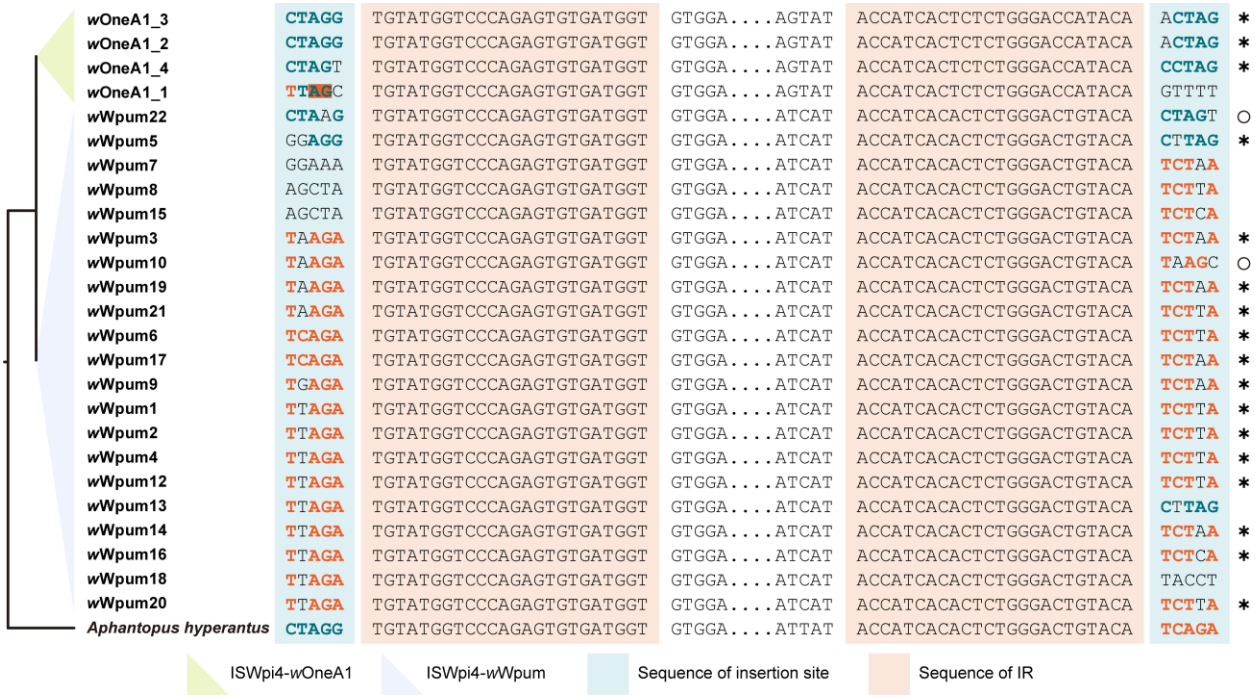
A: Relative divergence time estimation of homologous sequences of ISWpi19 constructed with nucleotide substitution model GTR+F+G4. We set the node shared by ISWpi19-*w*Wpum and ISWpi19-*w*CauA with arbitrary scaling of relative age at 1. *Wolbachia* strains in the tree covered three supergroups, labeled with diverse color stripes. **B:** Relative divergence time estimation of homologous sequences of ISWpi4 constructed with nucleotide substitution model HKY+F+G4. We set the timing when ISWpi4-wWpum11 diverged as unit 1. Green clade indicates close relationship of expanded ISWpi4 in *w*Wpum with homologous sequences from *w*OneA1 and insect species *Aphantopus hyperantus*. **C:** Relative divergence time estimation of homologous sequences of ISOt6 constructed with the same nucleotide substitution model as in (A). Orange clade indicates close relationship of ISOt6-*w*Wpum with homologous sequence from *Parachlamydia acanthamoebae*, and node shared by ISOt6-*w*Wpum and *w*Nik was set with the relative age of unit 1. Average divergence time of each main node in the three trees is

labeled with posterior probability in brackets.



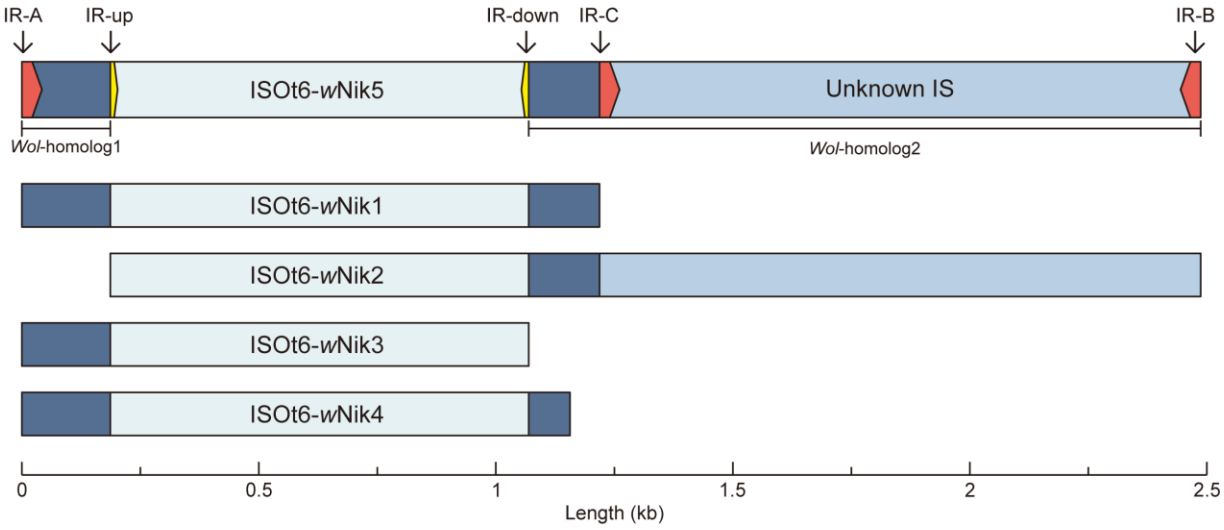
Supplementary Figure S4 Supergroup classification for *Wolbachia* strain *w*Nik

We determined supergroup classification of *w*Nik using MLST. Like *w*Wpum, *w*Nik belonged to supergroup A, but they were divided into two distinct clades. *w*Nik and *w*Wpum are both labeled with brown background and the clades they belong to are in red and blue, respectively.



Supplementary Figure S5 Comparisons of sequence features of 26 ISWpi4 elements

Typical ISWpi4 elements had 5 bp insertion site sequences and IRs in each terminal. In insertion site sequences, sequences similar to TCAGA are in orange and those similar to CTAGG are in green. Circles on the right of sequences represent that insertion site sequences at both ends of ISWpi4 elements are approximate direct repeats, and asterisks represent that insertion site sequences are approximate inverse repeats.



Supplementary Figure S6 Five ISOt6-*w*Nik copies with flanking homologous *Wolbachia* sequences in *w*Nik genome

ISOt6-*w*Nik5 had the longest flanking homologous sequences of *Wol*-homolog1 and *Wol*-homolog2, while other ISOt6 copies carried flanking homologous sequences with different lengths. In ISOt6-*w*Nik5, *Wol*-homolog2 harbored an unknown IS. Two groups of IRs were identified in the full-length sequence of "*Wol*-homolog1+ISOt6+*Wol*-homolog2", including two internal IRs

of ISOt6 17 bp long (yellow) and three IRs 44 bp long in the flanking regions (jacinth).

References

Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics*, **20**(3): 407–415.

Cerveau N, Leclercq S, Leroy E, Bouchon D, Cordaux R. 2011. Short- and long-term evolutionary dynamics of bacterial insertion sequences: insights from *Wolbachia* endosymbionts. *Genome Biology and Evolution*, **3**: 1175–1186.

Conner WR, Blaxter ML, Anfora G, Ometto L, Rota-Stabelli O, Turelli M. 2017. Genome comparisons indicate recent transfer of wRi-like Wolbachia between sister species *Drosophila* suzukii and *D. subpulchrella*. *Ecology and Evolution*, **7**(22): 9391–9404.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*, **27**(8): 1164–1165.

Ellegaard KM, Klasson L, Näslund K, Bourtzis K, Andersson SGE. 2013. Comparative genomics of *Wolbachia* and the bacterial species concept. *PLoS Genetics*, **9**(4): e1003381.

Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **11**(1): 119.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**(6): 587–589.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**(4): 772–780.

Klasson L, Westberg J, Sapountzis P, Näslund K, Lutnaes Y, Darby AC, et al. 2009. The mosaic genome structure of the *Wolbachia w*Ri strain infecting *Drosophila simulans*. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(14): 5725–5730.

Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, **13**(9): 2178–2189.

Miao YH, Xiao JH, Huang DW. 2020. Distribution and evolution of the bacteriophage WO and its antagonism with *Wolbachia*. *Frontiers in Microbiology*, **11**: 595629.

Price AL, Jones NC, Pevzner PA. 2005. *De novo* identification of repeat families in large genomes. *Bioinformatics*, **21 Suppl 1**: i351–i358.

Richardson MF, Weinert LA, Welch JJ, Linheiro RS, Magwire MM, Jiggins FM, et al. 2012. Population genomics of the *Wolbachia* endosymbiont in *Drosophila melanogaster*. *PLoS Genetics*, **8**(12): e1003129.

Siguier P, Filée J, Chandler M. 2006. Insertion sequences in prokaryotic genomes. *Current Opinion in Microbiology*, **9**(5): 526–531.

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, **4**(1): vey016.

Sutton ER, Harris SR, Parkhill J, Sinkins SP. 2014. Comparative genome analysis of *Wolbachia* strain *w*Au. *BMC Genomics*, **15**(1): 928.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, **56**(4): 564–577.

Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, et al. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. *PLoS Biology*, **2**(3): E69.