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.

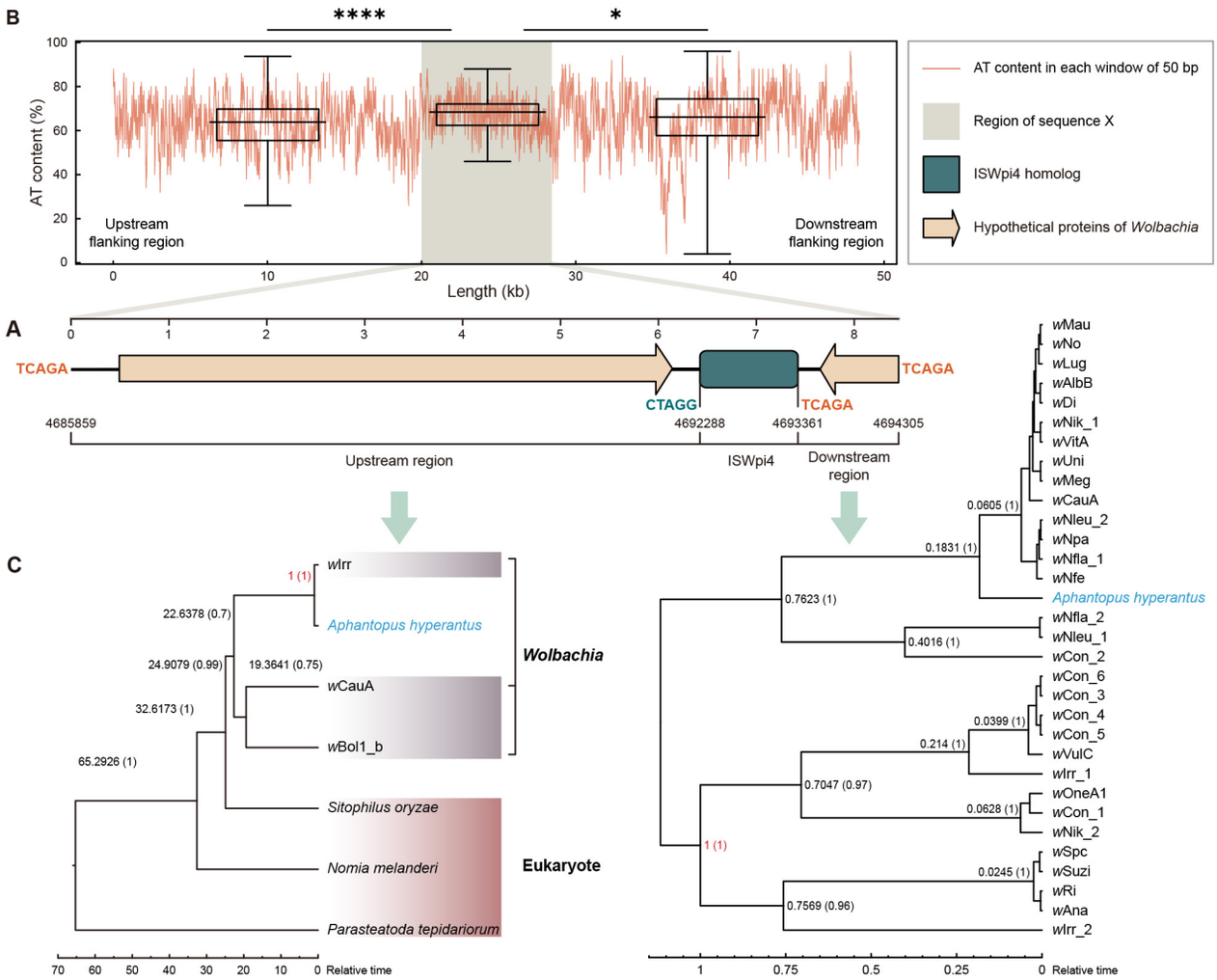

Figure 1 *Wolbachia* origin of sequence X in *A. hyperantus*

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.