Article

Title:

A co-speciation dilemma and a lifestyle transition with genomics consequences in *Wolbachia* of Neotropical *Drosophila*

Authors: Konstantinos Papachristos[1], Wolfgang J. Miller[2#], Lisa Klasson[1#]

Author affiliation: [1]Molecular evolution, Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden.

[2]Lab Genome Dynamics, Department Cell & Developmental Biology, Center for Anatomy and Cell Biology, Medical University of Vienna, Vienna, Austria.

[#]Corresponding authors: wolfgang.miller@meduniwien.ac.at and lisa.klasson@icm.uu.se

Emails:

konstantinos.papachristos@icm.uu.se

wolfgang.miller@meduniwien.ac.at

lisa.klasson@icm.uu.se

Keywords: *Wolbachia*, *Drosophila*, co-speciation, genome reduction, comparative genomics, symbiosis

Abstract

Long-term relationships between hosts and their bacterial endosymbionts can result in co-speciation. For obligate mutualistic associations, this occurs relatively frequently, but it is less common in parasitic or facultative relationships. Here, we have sequenced genomes of several very closely related *Wolbachia* strains, infecting fruit flies of the willistoni and saltans groups of Neotropical *Drosophila*, of which only one, the *w*Pau strain infecting *D. paulistorum*, is obligate for its host. Using these genomes, we constructed phylogenies and found congruence between *Wolbachia* and their hosts, except for *w*Pau and *D. paulistorum*, indicating mainly co-speciation between *Wolbachia* and willistoni and saltans group flies. However, the high similarity between *Wolbachia* strains implies low evolutionary rates and challenges the interpretation of our results as co-speciation. Additionally, contrary to the expectations of the genome reduction theory, we observed a genome size increase in the obligate *w*Pau strain, which is due to an ongoing expansion of an IS4 element. Finally, although *w*Pau has lost a large proportion of prophage WO genes, the *cif* genes are intact and nine genes from the WO-associated Undecim cluster are uniquely duplicated, suggesting they are under strong selection and potentially play a role in the obligate association.

Introduction

Bacterial symbionts are ubiquitous among animals but vary widely in phenotypic effects, level of dependency, and evolutionary persistence within host lineages. Mutualistic symbioses often involve tight associations with the host that may become obligate and thereby evolutionary persistent in a host lineage over time. In contrast, parasitic or pathogenic symbionts are usually facultative and evolutionarily short-lived unless they can manipulate their host to ensure their transmission across generations, by so-called reproductive parasitism (Zeng and Wiens 2021). Hence, the evolutionary persistence of a symbiont typically suggests that it is either an obligate mutualist or a reproductive parasite. The phenotypic effect, dependency, and persistence of a symbiotic association are thus interconnected and have been shown to influence how symbiont genomes evolve (McCutcheon and Moran 2012; Fisher et al. 2017). Numerous studies (Wernegreen 2015; McCutcheon and Moran 2012; Bobay and Ochman 2017) have demonstrated that obligate symbionts typically have smaller genomes, fewer pseudogenes, repeats, and mobile elements than facultative symbionts. The reduced genome size in obligate symbionts reflects their long-term reliance on the host, with the loss of non-essential gene functions and ongoing deletions facilitated by genetic drift. In contrast, facultative symbionts, which maintain a more transient association with their hosts, retain larger genomes that provide greater metabolic versatility and independence as well as more repeats and mobile elements.

Wolbachia is an intracellular symbiont that infects many different arthropod species as well as some nematodes (Kaur et al., 2021) and forms both facultative and obligate host associations. It is mainly vertically transmitted via the maternal line to the next generation (Russell et al. 2019) but is also frequently horizontally transmitted between host species (Vancaester and Blaxter 2023). *Wolbachia* can affect many aspects of its host's biology, including fecundity, immunity, and behaviour (Kaur et al. 2021). However, it is best known in arthropods for its ability to alter host reproduction. One such reproductive phenotype is cytoplasmic incompatibility (CI), where infected males are unable to produce offspring with uninfected females. The reproductive advantage thereby gained by infected females over uninfected females, can result in the spread of *Wolbachia* into a host population (Shropshire et al. 2020) as well as persistence in a population once the infection has reached a high frequency (Turelli and Hoffmann 1995). Even so, only a handful of examples of evolutionary persistent *Wolbachia* are currently known. Phylogenetic congruence, which is a hallmark of evolutionary persistence and co-speciation, has, for example, been observed between obligate mutualistic *Wolbachia* in filarial nematodes (Lefoulon et al. 2016) and bed bugs

(Balvín et al. 2018), but also in non-obligate associations between *Wolbachia* and *Nasonia* wasps where *Wolbachia* induces CI (Raychoudhury et al. 2008) and *Nomada* solitary bees (Gerth and Bleidorn 2016) where the phenotypic effect of *Wolbachia* is unknown. However, phylogenetic congruence between host mitochondrial DNA and *Wolbachia* is regularly observed for closely related host species, suggesting that the cytoplasmic co-inheritance of mitochondria and *Wolbachia* is frequently maintained at least over shorter times (Charlat et al. 2009; Ilinsky 2013; Zhang et al. 2013; Cariou et al. 2017; Bakovic et al. 2020).

Wolbachia genomes generally conform to the genome reduction theory. Facultative *Wolbachia* strains infecting various arthropods have genomes ranging between 1.1-1.8 Mb in size (Werren et al. 2008; Vancaester and Blaxter 2023) with a substantial fraction of mobile elements and other repeats and relatively many pseudogenes. Obligate mutualistic *Wolbachia* of nematodes have smaller genomes, ranging from 0.5-1.0 Mb, with few pseudogenes and a low fraction of mobile elements or repeats (Lefoulon et al. 2020; Dudzic et al. 2022). Additionally, prophage WO regions are present in nearly all *Wolbachia* genomes except in *Wolbachia* of nematodes, where they are absent or highly degraded (Gavotte et al. 2007; Bordenstein and Bordenstein 2022). This is important since prophage WO contains genes responsible for reproductive parasitism, such as *cifA* and *cifB* responsible for CI (Beckmann et al. 2017; LePage et al. 2017) and *wmk* involved in male-killing (Perlmutter et al. 2019), as well as the Octomom region affecting *Wolbachia* titer and virulence (Chrostek and Teixeira 2015; Duarte et al. 2021). Furthermore, the Eukaryotic Association Module (EAM) of prophage WO encodes several proteins suggested to play a role in *Wolbachia*-host interactions (Bordenstein and Bordenstein 2016).

In the Neotropical *Drosophila* of the willistoni and saltans groups, most tested species carry *Wolbachia* with high similarity to each other and the well-known *w*Au strain from *Drosophila simulans* (hereafter called *w*Au-like) (Miller and Riegler 2006). The high frequency of infections and similarity between the *Wolbachia* strains found in the two *Drosophila* groups indicates that their ancestor may have been infected with a *w*Au-like *Wolbachia* and that the two partners may have been co-speciating. However, this observation is puzzling, as co-speciation suggests evolutionary persistence and almost all the *Wolbachia* strains in willistoni and saltans group flies are believed to be facultative (Müller et al. 2012) and, in some cases, unable to induce CI and/or lack the *cif* genes (Martinez et al. 2015, Martinez et al. 2021). Only the *Wolbachia* strain *w*Pau infecting *D. paulistorum* from the willistoni group, a superspecies consisting of multiple reproductively isolated semispecies, is considered obligate mutualistic, as previous results have shown that all collected *D. paulistorum* flies are infected. Additionally, complete removal of *w*Pau by antibiotic treatment causes lethality and reducing the *w*Pau titer results in distorted ovary

development (Miller et al. 2010), thus affecting fecundity. Additionally, *w*Pau is likely involved in the ongoing host speciation of *D. paulistorum* as it contributes to assortative mating between different semispecies of *D. paulistorum* (Miller et al. 2010), affects pheromone profiles (Schneider et al. 2019) and contributes to the distinct gene expression pattern in heads and abdomens of three semispecies (Baião et al. 2019). *D. paulistorum* is also known to carry two distinct mitotypes, α and β, and multiple mitochondrial introgressions with related Neotropical *Drosophila* species likely occurred during divergence of its semispecies, as indicated by discordant mitochondrial and nuclear phylogenies (Baião et al. 2023). Thus, if a potential mitochondrial donor species carried *Wolbachia*, *w*Pau could have entered *D. paulistorum* via introgression.

The Neotropical *Drosophila* system thus offers the opportunity to investigate evolutionary persistence and co-speciation in non-obligate *Wolbachia* associations and study genomic differences between very closely related facultative and obligate *Wolbachia* strains. To do so, we sequenced and assembled *Wolbachia* genomes from five *Drosophila* species: *D. paulistorum, D. willistoni* and *D. tropicalis* from the willistoni group, and *D. prosaltans* and *D. septentriosaltans* from the saltans group and inferred whole genome phylogenies. We found congruence between four of five *Wolbachia* strains and host nuclear genes and full congruence between *Wolbachia* and host mitochondrial genomes. Additionally, we discovered that the genome of obligate *Wolbachia* strain *w*Pau was larger than closely related facultative strains due to the expansion of IS4 elements. However, the *w*Pau genome also had features commonly associated with ongoing genome reduction, like a lower coding percentage and more pseudogenes. Finally, despite a reduced prophage WO content, we found the WO-associated *cif* genes intact and a unique duplication of the Undecim cluster, suggesting they are under selection and thus potentially important for the symbiosis between *w*Pau and its host.

Results

We assembled three complete (*w*PauO11, *w*ProPE2 and *w*WilP98) and eleven draft genomes (Table 1, Table S1) from *Wolbachia* strains infecting three different species of the *Drosophila* willistoni group: seven *w*Pau strain variants from the Orinocan (OR) semispecies of *D. paulistorum* (as determined by the host phylogeny), three *w*Wil strain variants from *D. willistoni,* one *w*Tro strain variant from *D. tropicalis,* and two species of the *Drosophila* saltans group: two *w*Pro strains variants from *D. prosaltans* and one *w*Spt strain variant from *D. septentriosaltans* (Table S2). We refer to each strain variant by the *Wolbachia* strain name followed by the *Drosophila* line it was sequenced from, i.e., *w*PauO11 is the *w*Pau strain variant from the *D. paulistorum* line O11.

The three complete genomes sequenced in this study, *w*PauO11, *w*WilP98 and *w*ProPE2, were all similar in genome size and content to each other and to previously sequenced genomes of facultative *Wolbachia* strains (Table 1).

Co-speciation between *Wolbachia* and their Neotropical *Drosophila* hosts

To determine if *Wolbachia* co-speciated with their willistoni and saltans group hosts, we started by producing a phylogeny of our new genomes and other closely related *Wolbachia* genomes. To achieve that, we downloaded all publicly available *Wolbachia* genomes and inferred a phylogeny using the five Multi-Locus Strain-Typing (MLST) genes (Baldo et al. 2006) and the *wsp* gene. We determined 23 of the 1613 downloaded genomes to be closely related to ours (Table S3) and clustered their proteomes with the proteomes of our 14 genomes. The resulting 620 clusters with single-copy orthologous genes, all without detectable recombination, were used to produce a phylogeny.

The concatenated gene alignment contained only 0.55% parsimoniously informative sites (3,088 of 562,755 sites), showing the very close relationship between all 37 strains. Despite this, the inferred phylogeny had high support for a clade including the five *Wolbachia* strains from the willistoni and saltans *Drosophila* groups, *w*Tro, *w*Wil, *w*Pau, *w*Pro and *w*Spt (Figure 1). This clade, which also contained *Wolbachia* strain *w*Au from the old-world host *D. simulans*, further included *Wolbachia* strains *w*Inn, *w*Bor and *w*Inc, which infect *Drosophila* species of various other new-world species groups. Thus, from here on, these nine highly similar *Wolbachia* strains are what we refer to as *w*Au-like, and the strains infecting willistoni and saltans group flies are referred to as *w*Au-like$_{ws}$.

The concatenated gene alignment of these nine *w*Au-like genomes contained only 0.3% informative sites and 13.2% of the genes had no informative sites at all. Even so, inside the *w*Au-like clade, all strain and most strain variant relationships, except the position of *w*Pau (bootstrap 62), were highly supported. Thus, to evaluate the position of *w*Pau, we

149 calculated the gene (gCF) and site concordance factors (sCF) for the internal branches of
150 the subtree that only included one representative genome (*w*PauO11, *w*WilP98, *w*Tro,
151 *w*ProPE2 and *w*SptPLR) of each *w*Au-like$_{ws}$ strain, which are the focus of the co-speciation
152 analysis. The concordance factors represent the percentage of gene trees, gCF, or
153 informative sites, sCF, that support a split in the tree over all other possible splits. After
154 removing single-copy genes without any informative sites, we found that almost all
155 informative sites (99.6-100%) place *w*PauO11 between the *w*Wil/*w*Tro and *w*Pro/*w*Spt
156 clades and most of the genes (76.1-81.2%) also agree with this placement (Figure S1).
157 Thus, the position of *w*PauO11 in our inferred phylogeny had strong support from informative
158 sites but a bit less from individual genes. Furthermore, the low bootstrap value for the split
159 between *w*Pau and other strains only occurred when the *w*Inc strain of *D. incompta* was
160 included in our phylogenetic analysis. When *w*Inc was excluded, the *w*Pau strain variants
161 were in the same position in the tree but with high bootstrap support (100) (Figure S2).
162 Although we were unable to find the underlying reason for the conflicting signal, our
163 analyses still indicate that the position of *w*Pau is reliable.
164 When comparing our inferred *w*Au-like strain phylogeny to the nuclear phylogeny of
165 the willistoni and saltans group hosts (Baiaõ et al. 2023), we saw that the trees were
166 congruent except for the position of *w*Pau and *D. paulistorum* (Figure 2A). Rather than being
167 positioned inside the clade of the other two willistoni group strains, *w*Wil and *w*Tro, *w*Pau is
168 positioned between the willistoni and saltans group (*w*Pro and *w*Spt) strains. However, full
169 congruence was observed when comparing the *Wolbachia* tree with the phylogeny of
170 mitochondrial genomes from the same *Drosophila* species (Figure 2B) (Baiaõ et al. 2023).
171 To test if our *Wolbachia* data fit better with the host nuclear or mitochondrial topologies, we
172 calculated the log-likelihood difference per gene between topologies congruent with either
173 the mitochondrial or the nuclear phylogeny. We found that most *Wolbachia* genes strongly
174 supported the mitochondrial-like topology, and a few genes weakly supported the nuclear-
175 like topology (Figure S3). Overall, our results indicate that the *w*Au-like$_{ws}$ *Wolbachia* strains
176 are highly congruent with the host mitochondrial phylogeny, and partly congruent with the
177 nuclear phylogeny of the hosts.
178 For co-speciation to be true, the divergence times of the hosts and *Wolbachia* strains
179 should match (referred to as temporal congruence). To get an idea about temporal
180 congruence, we calculated relative pairwise synonymous substitution rates of *w*Au-like$_{ws}$
181 *Wolbachia* and host genes (Table S4) and compared them to previously described cases of
182 co-speciation (*Nomada* and *Nasonia*) (Raychoudhury et al. 2008, Gerth and Bleidorn 2016)
183 and horizontal transmission (*D. suzukii* subgroup) (Conner et al. 2017) of *Wolbachia*.
184 Furthermore, we calculated mutation rates for *w*Au-like$_{ws}$ *Wolbachia* as well as for *Wolbachia*

strains in *Nomada, Nasonia* and the *D. suzukii* subgroup, by dividing the pairwise synonymous substitutions rate by the divergence times of their hosts. The relative substitution rates had a broad range from 0.0003 for *Wolbachia* in the *D. suzukii* subgroup, where *Wolbachia* was suggested to be horizontally transmitted, to 0.31 for *Wolbachia* of *Nasonia,* which was suggested to be co-speciating. *Wolbachia* of *Nomada* had ratios between 0.015-0.15 and *w*Au-like$_{ws}$ *Wolbachia* between 0.0025 to 0.019, i.e., one to two orders of magnitude lower than in *Nasonia* and some *Nomada* pairs where *Wolbachia* is suggested to co-speciate with the hosts, but one to two orders of magnitude higher than in the *D. suzukii* subgroup where *Wolbachia* was suggested to be horizontally transmitted. The estimated mutation rates were $9x10^{-9}$ for *Wolbachia* in *Nasonia,* $4x10^{-10}$ to $1x10^{-9}$ in *Nomada Wolbachia* and $7x10^{-12}$ in *Wolbachia* from the *D. suzukii* subgroup, while the mutation rate for *w*Au-like$_{ws}$ *Wolbachia* was 1 $x10^{-10}$ to $5x10^{-10}$ synonymous substitutions per synonymous site per year (Table S4). Hence, the evolutionary rates of *w*Au-like$_{ws}$ did not match those from earlier suggested *Wolbachia* co-speciation or horizontal transmission examples.

Repeat expansion and rearrangements in the genome of obligate strain *w*Pau

Contrary to the expectations for an obligate strain, the *w*PauO11 genome was larger than all other *w*Au-like *Wolbachia* genomes (Table 1), had a lower coding content and 25-50% more repeats than both the genomes of *w*Au-like *Wolbachia* and two closely related facultative outgroup strains, *w*Mel and *w*Cin2USA1 (Table 1, Figure 3A). Specifically, we discovered a unique and extensive expansion of an IS4 element, which accounted for 9.9% of the total genome size of *w*PauO11 (Figure 3B).

Since repeats act as templates for genomic rearrangements via intragenomic homologous recombination, we examined gene order differences between our closely related genomes.From a whole genome alignment, we saw that the *w*PauO11 genome exhibited many changes in gene order when compared to other *w*Au-like genomes (Figure 4A). To quantify this, we identified pairwise syntenic blocks between all genomes and used this information to construct a phylogeny (Figure 4B). In the resulting tree, *w*PauO11 was on the longest branch, indicating that it has experienced more genomic rearrangements than the other *Wolbachia* genomes.

These findings highlight that the *w*PauO11 genome differs significantly from the genomes of other *w*Au-like$_{ws}$ strains in several ways, such as having a high proportion of non-coding DNA and repeats and frequent genomic rearrangements. Interestingly, a similar pattern, except for the large repeat expansion, is also seen in the genome of the *w*Au-like strain *w*Inn from *D. innubila.*

IS4 expansion is still ongoing in the *w*Pau genomes

Given the significant expansion of IS4 elements in *w*PauO11, we investigated whether these elements are still actively transposing in *w*Pau genomes. By identifying all complete and incomplete copies in the *w*PauO11 genome, we estimated that it originally had at least 96 complete IS4 elements (Figure S4). To quantify IS4 elements in all other *w*Pau genomes, we masked all but one copy in the *w*PauO11 genome, mapped the Illumina reads of each *w*Pau strain to it and calculated the coverage over the IS4 element relative to non-repetitive regions. Applying this approach to Illumina reads from *w*PauO11, we estimated that this genome harboured 98 IS4 copies, which closely matched our observation of 96 copies. When extended to other *w*Pau variants, the analysis revealed variability in copy numbers of the IS4 element across genomes (95-124 copies/genome) (Table S5). Additionally, by examining improper read pairs, we identified the positions of IS4 element insertions that occurred after the divergence of the *w*Pau strain variants (Table S6). To get a better resolution of the *w*Pau strain variant relationships and thus be able to calculate the number of insertions on each branch more correctly, we inferred a phylogeny based on whole genome SNPs between the *w*Pau variants (Figure S5). Using this tree topology, we identified seven IS4 element insertion events that occurred after the divergence of the *w*Pau strain variants (Table S7, Figure S6). Hence, both read mapping approaches indicated that the IS4 element is still actively transposing in these *w*Pau genomes.

Next, we approximated the rate and timing of IS4 element insertions in all complete genomes of the *w*Au-like clade. Using copy number counts, synteny comparisons and our phylogeny, we estimated that the genome of the ancestor of all *w*Au-like strains had four IS4 elements and the number of IS4 element insertions in all non-*w*Pau genomes to be 22 (Figure S6). Using branch lengths as an approximation of time, we calculated a rate of 6,571 insertions/(substitution/site) for the non-*w*Pau genomes. We inferred 88 IS4 element insertions on the branch leading up to the ancestor of the *w*Pau variants, by estimating a minimum of 93 IS4 element copies in the genome of the last common ancestor of *w*Pau. This gave us a rate of 123,422 insertions/(substitution/site). The seven IS4 element insertions in the *w*Pau strain variants gave us an estimated rate of 83,333 insertions/(substitution/site). The insertion rate of IS4 elements is thus 18.8 times greater in the branch leading to the *w*Pau ancestor and 12.7 times greater in the current *w*Pau than in the other analysed *Wolbachia* strains (Table S8). Our analyses thus showed that the IS4 element insertion rate was slightly higher in the *w*Pau genome in the past than it is today, but expansion is still ongoing with a relatively fast rate.

Major loss of prophage WO genes, but retention of CI genes and duplication of the Undecim cluster in *w*Pau

To explore which genes might be linked to phenotypic differences among *w*Au-like$_{ws}$ strains, we clustered the genomes based on the presence and absence of genes. Our analysis revealed that the genomes of *w*Pro and *w*Spt clustered together, as did those of *w*Wil, *w*Tro, and *w*Au, consistent with their phylogenetic relationships (Figure 1). However, the gene content of *w*Pau was notably distinct from other the *w*Au-like$_{ws}$ genomes (Figure 5).

We analysed the variation in gene content between the *w*Au-like$_{ws}$ strains further, by identifying which genes were uniquely present or absent in the three lineages, *w*Pau, *w*Wil/*w*Tro and *w*Pro/*w*Spt. Our results showed that *w*Pau (defined as *w*PauO11 and at least one *w*Pau draft genome) lacked many genes compared to the other *w*Au-like$_{ws}$ strains (Figure 5, Table S9), rather than having unique ones (Table 2, Table S10). Most missing genes were from prophage WO regions (Table 2), as suggested by the lower proportion of prophage WO sequences in the *w*PauO11 genome. Of the 39 missing prophage WO-associated genes, most encoded structural phage WO proteins, such as tail, head, or baseplate proteins, whereas only six belonged to the Eukaryotic Association Module (EAM).

Given the large number of missing prophage WO genes in *w*PauO11, we wanted to determine if they all came from the same prophage WO copy. The prophage WO regions identified in the complete *w*Au-like genomes were highly similar to either the WO-A or WO-B regions in *w*Mel (Wu et al. 2004) and are referred to here as WOa and WOb, respectively. By classifying the prophage WO proteins into WOa- and WOb-like, we found that WOa-like prophages are generally more degraded than WOb-like prophages in all genomes, with the Lysis, Tail, and Tail fiber modules of the former entirely missing in all *w*Au-like genomes as well as in *w*Mel and *w*Cin2USA1 (Figure 6). However, in *w*PauO11, the WOa-like prophage was almost completely absent, except for a few genes within the Replication and Repair module. The WOb-like prophage in *w*Pau also displayed a reduced gene content, particularly in the Connector/Baseplate module. Conversely, the *w*Inn genome, which also has a low proportion of prophage WO, retained all Baseplate/Connector genes in WOa but instead lost them in WOb. The presence of genes associated with both WOa and WOb in *w*Pau and other complete *w*Au-like genomes suggests that intact copies were likely present in their common ancestor. Thus, *w*Pau has predominantly lost genes from the WOa copy, but both prophage WO copies have experienced gene loss.

Apart from prophage WO genes, nine ankyrin repeat domain-containing genes (ANKs) and nine hypothetical proteins were absent from all *w*Pau strain variants but present in the *w*Wil/*w*Tro and *w*Pro/*w*Spt lineages. Of the remaining 11 protein clusters absent in all

288 *w*Pau, functions varied and included two non-WO phage proteins, a Fic protein, Type IV
289 secretion system protein VirB6, Malonyl-CoA decarboxylase, Cytochrome d ubiquinol
290 oxidase, subunit I and Exopolysaccharide synthesis ExoD-related protein. All missing protein
291 clusters in *w*Pau were found in *w*Au-like genomes from one or several of the non-*w*Au-like$_{ws}$
292 strains, indicating that they have likely been lost in *w*Pau rather than gained in the other
293 *w*Au-like$_{ws}$ lineages.
294 Much fewer protein clusters were uniquely absent in the other two *w*Au-like$_{ws}$
295 lineages. Twelve clusters were missing from the *w*Pro and *w*Spt genomes, and 20 were
296 absent in *w*Wil and *w*Tro (Table 2). Noteworthy, the *w*Wil/*w*Tro lineage lacks the prophage
297 WO-associated genes responsible for CI, *cifA* and *cifB*, as previously observed by Martinez
298 et al. 2021. Additionally, we found that *cifB* contains a frameshift mutation in the *w*Spt
299 genome, likely rendering it non-functional (Figure S7, Table S11). Furthermore, the *w*Bor
300 and *w*Inn strains also have frameshift mutations in *cifB* as well as non-sense mutations in
301 *cifA*, while *w*Inc has a non-sense mutation in *cifB* but an intact *cifA*. Thus, the only *w*Au-like
302 strains that potentially can induce CI are *w*Pro and *w*Pau. Since highly similar Type I *cif*
303 genes are found in all *w*Au-like genomes except in one clade (*w*Tro/*w*Wil/*w*Au), the common
304 ancestor of *w*Au-likes likely carried Type I *cif* genes. Hence, our results indicate that the *cif*
305 genes have been lost or pseudogenized independently multiple times in the *w*Au-like clade.
306 Only six protein-coding genes, encoding four hypothetical proteins, one ANK and a
307 ComF family protein, were present in *w*PauO11 (and at least one other draft *w*Pau genome)
308 but absent in the other *w*Au-like$_{ws}$ strains (Table 2). Potentially, one of the hypothetical
309 proteins and the ANK gene were acquired by *w*Pau, since no homologs were found in any
310 other *w*Au-like genome. Additionally, one of the hypothetical proteins is homologous to
311 WD0462 in *w*Mel, which correlated with CI induction in another set of *Wolbachia* strains
312 (Baião et al. 2021). However, since the gene is pseudogenized in *w*Pro, a strain known to
313 induce CI, the correlation was not seen among our strains. Finally, the ComF family protein,
314 with a potential role in DNA uptake, is pseudogenized in other *w*Au-like$_{ws}$ genomes.
315 More genes are uniquely present in the *w*Wil/*w*Tro and the *w*Pro/*w*Spt lineages. For
316 example, homologs of WD0463, the neighbouring gene of WD0462, are only intact in
317 *w*Pro/*w*Spt. This genomic region showed additional variability, with *w*Pau and *w*Pro carrying
318 an extra gene between WD0462 and WD0463 (Figure S8). Even though the function of
319 either of these genes is unknown, the region is highly variable between *Wolbachia* genomes
320 (Baião et al. 2021), even among these very close relatives. Additionally, we noted that six
321 genes in *w*Pro and five in *w*Spt from the Dozen Island, a genomic region likely stemming
322 from an integrated plasmid (Baião et al. 2021, Martinez et al. 2022), were not present in

323 *w*Pau or *w*Wil/*w*Tro. An additional four genes of the Dozen Island were found in *w*Pro/*w*Spt
324 and *w*Pau but lost from the wWil/wTro lineage.
325 Finally, although the loss of prophage WO genes appeared to be frequent in the
326 *w*Pau genome, we identified a large and unique duplication of a prophage WO-associated
327 region in *w*PauO11 containing ten of the eleven genes in the Undecim cluster (Bordenstein
328 and Bordenstein 2022) (Figure 7). To determine if the duplication was present in the other
329 *w*Pau strain variants, we estimated the copy number by mapping reads from those genomes
330 to the *w*PauO11 genome with one copy masked. Coverage over the Undecim cluster was
331 twice as large as coverage of non-repeated sequences for *w*PauFG111, *w*PauPOA1 and
332 *w*PauTP37 (Table S5), indicating duplication of Undecim in those genomes. For
333 *w*PauFG103, *w*PauFG295 and *w*PauRP, coverage over the Undecim cluster was more like
334 in non-repeated sequences. However, for those lines, the standard deviation of coverage for
335 non-repeats was as much as 30% of the mean, compared to 10% for the other lines. This
336 suggested that non-repeat coverage may be noisy for some samples, and we thus plotted
337 the coverage of reads from all strains around and over the Undecim cluster of wPauO11
338 (Figure S9). Since the coverage was approximately doubled compared to the surrounding
339 single-copy regions for all *w*Pau strain variants, the same genomic region is likely duplicated
340 in all sequenced *w*Pau genomes.

Discussion

Neotropical *Drosophila* species of the willistoni and saltans groups were previously shown to carry closely related *Wolbachia* strains with high similarity to the *w*Au strain from *D. simulans* (Miller and Riegler 2006). Using 14 new *Wolbachia* genomes from five Neotropical *Drosophila* species, we confirm that these strains are highly similar and belong to a monophyletic clade including strain *w*Au – i.e. the *w*Au-likes.

The *w*Au-like *Wolbachia* show patterns of co-speciation and horizontal transmission

To investigate the possibility of co-speciation between the *w*Au-like *Wolbachia* strains infecting willistoni and saltans group flies, *w*Au-like$_{ws}$ strains, and their hosts, we inferred a phylogeny and compared it to the nuclear and mitochondrial phylogenies of the hosts (Baião et al. 2023). We found that the phylogeny of *w*Au-like$_{ws}$ strains was congruent with the nuclear phylogeny of the host species, except for the *Wolbachia* strain *w*Pau and its host *D. paulistorum*. Complete congruence was, however, found between the topologies of the two maternally transmitted genetic entities, *Wolbachia* and mitochondria. Hence, based on phylogenetic congruence, it seems plausible that a *w*Au-like strain was present in the ancestor of the willistoni and saltans groups at their divergence ca. 20 million years ago (Powell et al. 2003) and that this strain has co-diverged with the current hosts ever since.

However, phylogenetic congruence can be achieved in the absence of co-speciation, a phenomenon referred to as pseudo-co-speciation, especially common if host switching preferentially occurs between closely related species (De Vienne et al. 2007). Thus, to infer true co-speciation, temporal congruence between the two compared entities should ideally be observed. However, since no reliable estimate of the mutation rate exists for any *Wolbachia* strain, and the inferred divergence dates of host species are not confirmed by fossil data, dating the nodes in either of our trees will be highly unreliable. Nevertheless, to attempt to investigate temporal congruence, we calculated the synonymous substitution rate of *w*Au-like$_{ws}$ strains relative to that of their hosts and compared that to previously reported cases of *Wolbachia*-host co-speciation. We found the relative rates in our dataset to be quite constant between different *Wolbachia*-host pairs, with *D. paulistorum*- *D. willistoni* and their *Wolbachia* having the highest ratio around 3-7 times higher than the others. Even so, all rates were in general between one and two orders of magnitude lower than in previous co-speciation cases (Raychoudhury et al. 2008; Gerth and Bleidorn 2016; Conner et al. 2017). Thus, if all co-speciation cases were true, the *w*Au-like$_{ws}$ strains must either have evolved much slower than the other co-speciating *Wolbachia* strains or the willistoni and saltans group flies must have evolved much faster than the hosts in previous studies. However, this

comparison is not conclusive as we don't know if the earlier reported cases represent true examples of co-speciation, since neither has proven temporal congruence. Additionally, mutation and substitution rates as well as generation times can vary widely between different organisms, even between very closely related bacterial and animal species.

Therefore, we also calculated mutation rates of *Wolbachia* based on the synonymous substitution rates and the host divergence times. Using this method, we got mutation rates between 10^{-10}–10^{-8} synonymous substitutions per synonymous site per year, which is lower than in other bacteria (Gibson and Eyre-Walker 2019). Consequently, for our calculated rates to match the generational mutation rates of bacteria (ranging between 10^{-10}–10^{-8} synonymous substitutions per synonymous site per generation (Lynch et al. 2023)), all *Wolbachia* strains*,* except the ones in *Nasonia,* must have a generation time of almost a year. Since such a generation time is even longer than the hosts', our estimated mutation rates are likely incorrect. There are indeed many potential sources of error for our mutation rate estimates, for example, synonymous substitutions could be under strong selection, or the recombination rate could be extremely high and eliminate any diversity between strains, both of which could yield underestimated mutation rates. Additionally, the host divergence times could be wrong, which, depending on if they are over- or under-estimated, could result in either lower or higher mutation rates. At this point, we can't completely rule out any of these factors, although selection seems unlikely given that the effective population size of *Wolbachia* is probably low, and it is hard to understand how recombination between the *w*Au-like$_{ws}$ strains could be so rampant in the absence of horizontal transmission. Even though there is a good chance that the divergence times of the hosts are wrong, they would have to be off by orders of magnitude, which is less likely. Alternatively, for co-speciation to be believable, *Wolbachia* must have a very low mutation rate or a long generation time, or a combination thereof. If not, neither the *w*Au-like$_{ws}$ strains nor the *Wolbachia* strains from *Nomada* and *Nasonia* have co-speciated with their hosts, but may instead have been horizontally transmitted into their hosts via introgression or alternative mechanisms, as suggested in other similar systems (Turelli et al. 2018; Cooper et al. 2019).

It is unlikely that *Wolbachia* could have been recently transmitted between the relatively distantly related species of the willistoni and saltans group via introgression, but full congruence between mitochondria and *Wolbachia* suggests that introgressions have facilitated at least some horizontal transmission. As previously reported, *D. paulistorum* has two mitotypes, α and β, which were probably both introgressed from an unknown species within the willistoni group (Baião et al. 2023). Hence, it is likely that *w*Pau was introgressed into *D. paulistorum* from a willistoni group species together with one of the mitotypes, potentially β since it is most common among *D. paulistorum* semispecies and all *w*Pau

strains in this study are from *D. paulistorum* lines carrying the β mitotype. As *w*Pau carries complete and thus likely functional *cif* genes, the spread of *w*Pau and the replacement of the ancestral *D. paulistorum* mitochondrion could potentially have been driven by CI. However, to fully answer this question, data from additional *D. paulistorum* semispecies, especially the ones carrying the α-mitotype, are needed.

Additionally, at least four horizontal transmission events of *w*Au-like *Wolbachia, w*Inn, *w*Bor, *w*Inc, and *w*Au, into distantly related *Drosophila* species have likely occurred relatively recently, as previously suggested (Ballard 2004; Miller and Riegler 2006; Sheeley and McAllister 2009; Wallau et al. 2016), but through mechanisms other than introgression. We can only speculate about how these *w*Au-like *Wolbachia* strains infected and spread in their new hosts, but CI is not a likely mechanism for the spread since none of them carry what appear to be functional *cif* genes. The well-studied strain *w*Au has long been known to not induce CI in its native host, *D. simulans* (Hoffmann et al. 1996). However, it does provide high antiviral protection in its host (Martinez et al., 2015), a potentially beneficial phenotype that might have facilitated its spread. Nothing is currently known about the phenotypic effects *w*Inc has on its host, but both *w*Inn and *w*Bor cause male-killing, which could have contributed to their spread. Furthermore, *w*Inn was also seen to protect the host against RNA viruses (Unckless and Jaenike 2012). It is also possible that *w*Au-like strains provide yet unknown beneficial host effects.

In conclusion, more research is required to determine whether *Wolbachia* switches hosts so frequently that pseudo-co-speciation is observed independently in multiple clades and to identify mechanisms that facilitate the transfer and spread of *Wolbachia* infections.

The genome of obligate *Wolbachia* strain *w*Pau is expanded and has some genomic features similar to recently host-associated symbionts

When free-living bacteria become host-associated, many of their genes become superfluous in the new nutrient-rich, stable environment of the host and are hence lost or pseudogenized due to lack of selection (McCutcheon and Moran 2012). In addition, when symbionts transition from free-living to host-associated, they experience a reduction in effective population size that increases the effect of genetic drift (Moran and Plague 2004, Wernegreen 2015). Consequently, their genomes tend to accumulate slightly deleterious mutations, including many IS element insertions (Bobay and Ochman 2017). However, over time, often after becoming obligate for their hosts, symbiont genomes generally shrink further in size and lose repeated sequences and pseudogenes (McCutcheon and Moran 2012, Bobay and Ochman 2017), leading to a conserved gene order. This stage of the genome reduction process can, for example, be observed in obligate mutualistic *Wolbachia*

447 of nematodes that have small genomes (Werren et al. 2008, Dudzic et al. 2022) with few
448 repeats and mobile elements (Lefoulon et al. 2020) compared to most facultative *Wolbachia*
449 strains.

450 To our knowledge, *w*Pau is the only *Wolbachia* strain found in the willistoni and
451 saltans group of flies which is obligate for the host (Miller et al. 2010). Hence, *w*Pau is
452 expected to have a smaller genome than its close facultative relatives. Contrary to the
453 expectation, we found that *w*PauO11 has the largest genome among the *w*Au-like
454 *Wolbachia* strains with complete genomes. Genome size increase in *Wolbachia* is generally
455 associated with a higher proportion of prophage WO (Kampfraath et al. 2019; Lefoulon et al.
456 2020; Wolfe et al. 2021; Vancaester and Blaxter 2023), but in *w*Pau, we rather found a
457 reduction of prophage WO content. Instead, we attribute the increase in genome size in
458 *w*Pau to a still ongoing expansion of an IS4 element that makes up almost 10% of the
459 genome and has greatly affected the gene order. The extensive IS4 element proliferation,
460 high rate of genomic rearrangements and reduced coding percentage in the *w*Pau genome
461 are genomic features that are more common in recently host-associated facultative
462 symbionts than evolutionary persistent obligate symbionts. All of them, however, suggest a
463 large influence of genetic drift, potentially caused by a reduction in effective population size.
464 We can think of at least three possible reasons why the effective population size of *w*Pau
465 could have decreased compared to its close relatives.

466 Firstly, the horizontal transmission of *w*Pau might have reduced the effective
467 population size since only a few *Wolbachia* cells likely established the infection in *D.*
468 *paulistorum* initially. While we don't see an expansion of IS elements in other *w*Au-like
469 strains with complete genomes that have likely shifted host recently, i.e. *w*Inn and *w*Au, the
470 *w*Inn genome shares other features with the *w*Pau genome, including many rearrangements
471 and low coding density. Possibly, *w*Inn had more repeats but lost them over time due to
472 deletions. Although these similarities with *w*Inn might indicate that horizontal transmission
473 could initially cause a higher effect of genetic drift, we do not believe this to be the main
474 reason for the sustained high levels of genetic drift in *w*Pau.

475 Secondly, the transition from facultative to obligate host association could have
476 yielded a reduction in the effective population size of *w*Pau. Several similar genome features
477 can be found in the genome of the mutualistic *Wolbachia* strain *w*Cle infecting bedbugs
478 (Hosokawa et al. 2010), including an expansion of IS elements (with over 100 copies of IS5)
479 and reduced prophage WO content (Lefoulon et al. 2020). Potential IS element expansions
480 are also recorded in other obligate symbiont genomes, for example, the co-obligate *Serratia*
481 *symbiotica* strain Ct from the aphid *Cinara tujafilina* (Manzano-Marín and Latorre 2014),
482 where several IS element families have expanded (Renoz et al. 2021) and some species of

483 the genus *Sodalis* (Oakeson et al. 2014; Koga and Moran 2014; Ankrah et al. 2018;
484 Manzano-Marín et al. 2018; Garber et al. 2021). However, since no genome sequences of
485 closely related facultative strains are available in any of these cases, the IS element
486 expansion can't be linked to a transition to obligate host association.

487 Thirdly, a reduction in the effective population size of *w*Pau could be a consequence
488 of a reduction in the effective population size of its host, which might be linked to the fact that
489 *D. paulistorum* is a superspecies consisting of multiple fully or partly reproductively isolated
490 semispecies. Currently, no data about the effective population sizes of any species in the
491 willistoni or saltans groups exist, however, indirect evidence indicates a lower effective
492 population size in *D. paulistorum* compared to other willistoni and saltans group species.
493 Large eukaryotic genomes with a high proportion of transposable elements often occur in
494 species with low effective population sizes (Lynch and Conery 2003), due to the high
495 influence of genetic drift that allows mobile elements to accumulate. Since *D. paulistorum*
496 has the largest genome with the highest repetitive content of the sequenced willistoni group
497 species (Kim et al. 2021, Gebert et al. 2024), we might speculate that *D. paulistorum* has a
498 lower effective population size than other willistoni and saltans group species. If so, this
499 could explain not only the high rate of IS4 element insertions before the divergence of the
500 *w*Pau strain variants but also the continued high rate that seems to be ongoing in *w*Pau even
501 after its likely introgression and transition from facultative to obligate.

502 Prophage WO genes may be important for the obligate symbiosis of *w*Pau

503 The WO prophage is known to play an important role in facultative *Wolbachia* host
504 associations, given that it carries genes directly responsible for reproductive manipulation
505 (Beckmann et al. 2017; LePage et al. 2017; Perlmutter et al. 2019; Fricke and Lindsey
506 2024), infection titer and virulence (Chrostek and Teixeira 2015; Duarte et al. 2021). For the
507 obligate *w*Pau, we demonstrate loss of prophage WO genes compared to closely related
508 facultative strains. Specifically, *w*Pau has lost almost an entire copy (WOa) and some parts
509 of the other (WOb), suggesting a general lack of selection on WO-associated genes rather
510 than any specific WO prophage region. In contrast, nine of the eleven genes of the prophage
511 WO-associated Undecim cluster are instead duplicated in *w*PauO11 as well as the other
512 *w*Pau strain variants. Although the functions of the Undecim cluster genes are currently
513 unknown, they are thought to play a role in *Wolbachia*-host interactions, as they, among
514 other functions, encode proteins potentially involved in exopolysaccharide and/or
515 lipopolysaccharide biosynthesis (Bordenstein and Bordenstein 2022). Previous work has
516 also shown that the Undecim cluster has been horizontally transferred between several
517 unrelated insect endosymbionts and that the genes are constitutively highly expressed in

518 *Wolbachia* strain *w*Mel during the whole life cycle of *D. melanogaster* (Gutzwiller et al. 2015).
519 The fact that this prophage WO-associated set of genes is duplicated in *w*Pau, while many
520 other prophage WO genes are deleted, suggests that the region could encode proteins
521 potentially important for the obligate symbiosis between *w*Pau and its host. If true, prophage
522 WO might thus contribute genes for the persistence of a *Wolbachia* infection, even in the
523 absence of CI or other reproductive manipulations.

524 Conclusions

525 We conclude that the ancestor of Neotropical *Drosophila* in the willistoni and saltans groups
526 may have been infected with a *w*Au-like *Wolbachia* and co-speciated with their hosts, except
527 for the *w*Pau strain in *D. paulistorum*, which was likely introgression from another species
528 within the willistoni group together with one of the two mitotypes. However, such an
529 interpretation suggests that *Wolbachia* must have a very low mutation rate and/or a long
530 generation time. Further research is needed to determine if this is the case.
531 We also found that contrary to expectations, the genome of the obligate *w*Pau strain
532 was larger than its facultative relatives due to the ongoing proliferation of an IS4 element,
533 typically associated with genomes of recently host-associated symbionts. A reduced coding
534 percentage and more pseudogenes are also hallmarks of the early stages of genome
535 reduction, all of which are likely due to a reduction in the effective population size.
536 In general, the effective population size can be reduced multiple times after a
537 bacterium becomes host-associated. Thus, the genome reduction process might take place
538 in a punctuated manner, as episodes of effective population size reduction can temporarily
539 lead to an increased genome size via repeat expansion, and a decrease in the proportion of
540 coding content, only to be reversed by homologous recombination over such repeats over
541 time. Consequently, only when mobile elements and repeats are lost from a genome
542 altogether, can this cycle be broken.

Materials and Methods

DNA Extraction and Sequencing

Drosophila lines used in this study and the sequencing data produced from each of them are summarized in (Table S2). For *D. paulistorum* O11, total DNA for Oxford Nanopore (ONT) sequencing was extracted from many adult flies of both sexes using the MagAttract HMW DNA Kit (Qiagen) to obtain high molecular weight DNA. For Illumina sequencing, total DNA was obtained as described in Baiaõ et al 2023. For *Wolbachia* strains *w*WilP98, *w*ProPE2, *w*PauFG111, *w*PauPOA1 and *w*PauTP37, DNA samples used for PacBio sequencing were enriched for *Wolbachia* using a differential centrifugation and filtration procedure to obtain a *Wolbachia* cell pellet, which was then subjected to whole genome amplification using the Repli-g midi kit (Qiagen) as described in (Ellegaard et al., 2013). For *w*WilP98 and *w*ProPE2, Illumina sequences enriched for *Wolbachia* were obtained using the same protocol.

The ONT library was produced and sequenced using an R9.4 PromethION flow cell at the Uppsala Genome Center, Uppsala, Sweden. Basecalling was performed with Guppy 4.3.4 and the HAC model. Illumina TruSeq libraries were produced and sequenced at the Uppsala SNP and Seq platform on an Illumina 2500 HiSeq machine generating 2x125 bp reads. The Illumina reads were quality filtered and trimmed using Trimmomatic-0.30 with parameters –phred33 ILLUMINACLIP:illumina_adapter.fasta:2:40:15 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:95 (Bolger et al., 2014) and error-corrected using SPAdes-3.5.0 –sc –only-errorcorrection (Bankevich et al., 2012). For PacBio, 5 kb SMRTbell libraries were created. Each library was run with P6-C4 chemistry in one SMRT cell on an RSII PacBio instrument at Uppsala Genome Center.

Genome Assemblies

Wolbachia genomes were assembled using several types of data and pipelines. For *w*PauO11, reads from total DNA sequencing (i.e. including the *Drosophila* host) were filtered using Filt-long v0.2.1 with quality priority and the parameters "--target_bases 12500000000 --min_length 1000 --mean_q_weight 10" and assembled with Nextdenovo v2.5.0 using correction options: read_cutoff = 1k, genome_size = 250m, sort_options = -m 7g -t 2, minimap2_options_raw = -t 3, pa_correction = 17, correction_options = -p 2 and assemble options: minimap2_options_cns = -t 3, nextgraph_options = -a 1. The produced assembly contained only one *Wolbachia* contig that was subsequently polished with Illumina reads Pilon v1.24 (Walker et al. 2014) with the --fix indels parameter and circularized. The genomes of the two strains *w*WilP98 and *w*ProPE2 were assembled using Illumina and PacBio reads from amplified DNA enriched for *Wolbachia*. The Illumina reads were assembled using Spades v3.15.5 (Bankevich et al. 2012) with kmer sizes 97,103,107,113

578 and the PacBio reads were used for scaffolding with the --pacbio parameter. Illumina and
579 PacBio reads were mapped against the assemblies using BWA-mem (Li and Durbin 2009)
580 with default parameters for Illumina reads and using PacBio settings for PacBio reads to
581 check the correctness of the genome assembly. Overlapping contigs were merged in
582 Consed (Gordon et al. 1998). For all other assemblies, BWA-mem v0.7.17-r1198-dirty with
583 default parameters was used to map Illumina reads on the *w*PauO11, *w*ProPE and *w*WilP98
584 genomes and a set of concatenated *Wolbachia* genomes representing the *Wolbachia*
585 diversity (81 genomes from supergroups A, B, C, D, E, F, L and M, Table S12). Mapped
586 reads were filtered using samtools v.1.13 (Danecek et al. 2021) view with the -f 1 -F 12
587 parameters to include only pairs where both were mapped. Filtered reads were then
588 subsampled to approximately 100X with seqtk v1.3-r117-dirty sample -s 100 before
589 assembly. The reads were then assembled using Spades v3.15.5 with kmer sizes
590 97,103,107,113, and the *w*PauFG111, *w*PauPOA1 and *w*PauTP37 assemblies were
591 scaffolded with PacBio reads using the flag -pacbio.
592 The Illumina assemblies were then filtered by removing contigs that deviated from
593 the mean sample distribution in kmer coverage and GC content. The kmer coverage of the
594 contigs was determined from the fasta headers in the output from Spades. In cases where
595 kmer coverage had a bimodal distribution, the minimum between the two peaks was
596 identified using the quantmod package in R (Ryan and Ulrich 2022) and used as threshold.
597 Contigs with kmer coverage below this threshold were blasted against the assembled
598 contigs and were removed if the sequence existed. This procedure ensured that redundant
599 low coverage contigs were removed. Remaining low coverage contigs were blasted against
600 additional *Wolbachia* genomes and proteins. Contigs with a blast alignment length longer
601 than 100 aa and percent identity over 60%, or blast alignment length longer than 200
602 nucleotides and percent identity over 70%, were kept. Any contig with an extreme GC value
603 (outside the 99% interval of a normal distribution of GC) was removed. Finally, all remaining
604 contigs were searched with Kraken against a custom database (containing a selection of
605 archaea, bacteria, plasmids, viruses, human, *D. melanogaster* and *D. willistoni* sequences)
606 and the contigs with a match to *Drosophila* were removed. All assemblies were polished
607 once using Pilon with the Illumina reads used for assembly.

608 Genome Annotation

609 For the complete genomes, *w*PauO11, *w*ProPE2 and *w*WilP98, Prokka v1.14.6 (Seemann
610 2014) with the parameters –genus Wolbachia –compliant –rfam --addgenes --usegenus was
611 used for annotation. Next, Pseudofinder v1.1.0 (Syberg-Olsen et al. 2022) with parameters -
612 ce and -db with a database containing all *Wolbachia* proteins from NCBI (DATE), was used

613 to mark potential pseudogenes. Additionally, hmmsearch as implemented in pfam_scan.pl
614 was used for domain prediction with the PFAM database (Bateman et al. 2002). The Illumina
615 reads used for the assemblies were mapped back to each genome with BWA-mem to verify
616 frameshifts and nonsense mutations. All marked pseudogenes were inspected manually and
617 annotated as pseudogenes if the start or stop codons were missing, they contained
618 nonsense or frameshift mutations or were truncated by, for example, an IS element
619 insertions.
620 The draft assemblies were annotated the same way, except that 1) Pseudofinder was
621 run with length_pseudo set at 0.8 and the protein database only included *Wolbachia*
622 genomes *w*Au, *w*Ha, *w*Ma, *w*No, *w*San, *w*Tei, *w*Yak, *w*PauO11, *w*ProPE2 and *w*WilP98 that
623 were manually annotated either in this or previous studies (Ellegaard et al. 2013; Baiaõ et al.
624 2021), 2) only pseudogene calls in CDS sequences were inspected manually, 3)
625 pseudogenes that we identified in the complete genomes (*w*PauO11, *w*ProPE2 and
626 *w*WilP98), but were not found by pseudofinder in the draft genomes, were transferred to the
627 draft genome annotation if they matched with high similarity and extensive coverage.
628 For identifying repeated sequences, nucmer from the MUMmer 4.0 package (Marçais et al.
629 2018) was used with flags --maxmatch --nosimplify to align each genome to itself. The
630 positions of non-overlapping repeat bases were then extracted using a perl script. ISEScan v
631 1.7.2.3 (Xie and Tang 2017) was used to identify IS elements. Bar plots of the percentage of
632 repeats and IS elements in each genome were made in R (R Core Team 2018; Wickham et
633 al. 2019). Phage WO regions were annotated by blastp searches to *Wolbachia* genomes
634 (*w*Mel, *w*Au, *w*Ha, *w*Ma, *w*No, *w*San, *w*Tei, *w*Yak) with previously identified phage WO
635 regions. All genomic regions that contained similarities to known phage WO genes were
636 then manually inspected, and borders were determined based on Bordenstein and
637 Bordenstein 2022.

638 MLST phylogeny of *Wolbachia* genomes

639 Genomes that contained the keyword “Wolbachia” in the taxon name, were downloaded
640 from NCBI using the command line tool “datasets” (https://github.com/ncbi/datasets) on May
641 the 6th 2022. tblastn of the Blast v2.7.1+ suite (Camacho et al. 2009) with the parameters --
642 max_hsps 1 --evalue 1e-10 was used to search for the five *w*Mel MLST proteins and
643 Wolbachia surface protein (Wsp) (GenBank accession numbers AAS13898.1, AAS14036.1,
644 AAS14201.1, AAS14415.1, AAS14719.1, AAS14884.1) in 1504 downloaded genomes.
645 Genomes with blast hits for all six genes were kept for phylogenetic inference. Nucleotide
646 sequences of the identified genes were extracted from each genome using samtools faidx,
647 and translated, aligned with MAFFT v6.850b using the linsi algorithm (Katoh and Toh 2008),

648 pruned to remove gap sites present in more than 50% of aligned sequences and then
649 backtranslated to nucleotides again. A phylogeny of the six genes from 1262 genomes was
650 then constructed using RaxML v8.2.9 (Stamatakis 2014) with parameters –f a –m
651 GTRGAMMA –x 12345 –p 12345 -# 100. The resulting phylogenetic tree as visualized in
652 Figure 1 and all other figures presenting phylogenetic trees (Fig. 2, Fig. 4B, Fig. S1, Fig. S2,
653 Fig S3, Fig. S5 and Fig. S6) were made using FigTree v1.4.4 (Rambaut 2009).

654 Protein clustering and strain phylogeny

655 At the time of downloading *Wolbachia* genomes, the ones from Vancaester and Blaxter 2023
656 were not yet available. Hence, to include them in our analysis, we downloaded the genomes
657 classified as supergroup A by the authors. We then annotated the genomes from all strains
658 classified as supergroup A *Wolbachia* in the MLST phylogeny and by Vancaester and
659 Blaxter 2023 as described above for our draft assemblies, used their proteomes for
660 clustering and inferred a phylogeny from single-copy genes. Clustering was done with
661 Orthofinder v2.5.4 (Emms and Kelly 2019) with –I 3 and filtered all-against-all blastp results
662 as input. Blastp hits between two matching proteins were filtered so that the size of one
663 should not be smaller than 60% of the size of the other, the e-value should be lower than 10^{-}
664 ^{5} and the alignment length should be at least 80% of the size of the smaller protein. The
665 nucleotide sequences of genes from orthogroups with single-copy proteins were extracted,
666 translated, aligned with mafft-linsi, pruned to remove gap sites present in more than 50% of
667 the aligned sequences and backtranslated to obtain codon-based nucleotide alignments.
668 To ensure that the genes used in the phylogeny did not show signs of recombination,
669 we ran Phipack (Bruen et al. 2006) with the -p 100000, -o, -w 50 on all alignments. Phipack
670 runs three tests to detect recombination, the Phi test, the max $\chi 2$ test and the NSS test and
671 at least two of the three recombination tests needed to have a p-value lower than 0.05, for a
672 gene to be considered as recombined. Single-copy gene alignments without recombination
673 were then concatenated with geneStitcher.py from Utensils
674 (https://github.com/ballesterus/Utensils/blob/master/geneStitcher.py).
675 IQ-TREE v2.2.0 (Minh et al. 2020) was then used for phylogenetic inference of the
676 species tree using the concatenated alignments of the single-copy genes, a partition per
677 gene, and the parameters -m MFP, -B 1000 -T 20 --seed 12345. Individual gene trees were
678 inferred using the same parameters. The phylogenetic trees were visualized using FigTree
679 v1.4.4. To test if the data fit better with the placement of *w*Pau according to the host nuclear
680 phylogeny rather than our strain phylogeny, we followed Shen et al. 2017 and used IQ-TREE
681 v1.6.12 with the same parameters as in the strain phylogeny together with the -z flag to add
682 the two topologies (shown in Figure S3) in Newick format and the -wsl and -wpl flags to get

683 the likelihood for each site and gene. The differences in the gene likelihoods between the
684 two topologies were calculated and plotted in R (R Core Team 2018; Wickham et al. 2019).
685 To measure the gene (gCFs) and site (sCFs) concordance factors, we used the same
686 pruned concatenated phylogeny as in the log-likelihood differences and we produced pruned
687 gene trees including the same strains with preserved branch lengths using a Python script
688 and the ete3 library (Huerta-Cepas et al. 2016). To calculate the gene and site concordance
689 factors, the pruned species tree was then provided to IQ-TREE with the flag -t, the fasta
690 alignments with -s, the partition file with -p, the gene trees were provided with --gcf and --scf
691 100 was used.

692 Substitution rates

693 Pairwise synonymous substitution rates were calculated with codeml from the PAML v4.8
694 package (Yang 2007) using alignments of the 620 single-copy *Wolbachia* genes from the
695 ortholog clustering, as well as the alignments of 692 BUSCO genes and 13 mitochondrial
696 genes from *Drosophila* (the same as in Baião et al 2013) generated the same way as the
697 *Wolbachia* alignments (see above).

698 Gene order phylogeny

699 Syntenic blocks between the complete genomes listed in Table 1 were obtained using
700 SynChro_linux (January 2015) (Drillon et al. 2014) with a delta value of 2. The output from
701 SynChro was then used for phylogenetic reconstruction with PhyChro (downloaded July
702 2023 from http://www.lcqb.upmc.fr/phychro2/) (Drillon et al. 2020).

703 IS4 element and Undecim cluster copy numbers in draft *w*Pau genomes

704 To estimate the number of IS4 elements and Undecim cluster copies in the draft *w*Pau
705 genomes, we first masked all copies of the IS4 element and one copy of the Undecim cluster
706 in the *w*PauO11 genome. To do that, we created an IS4 element consensus sequence by
707 blasting (blastn) one IS4 element from *w*PauO11 against the *w*PauO11 genome, extracting
708 the sequences for all matches, aligning them with MAFFT v7.490 with flags "--maxiterate
709 1000" and "--localpair" and building a consensus with Jalview v2.11.4.0 (Waterhouse et al.
710 2009). The consensus IS4 element sequence was then blasted against the *w*PauO11
711 genome using blastn and all hits were masked. The Undecim cluster copies were first
712 identified by homology to the Undecim cluster genes in the *w*Mel genome. One of the
713 Undecim cluster copies, at positions 898.960-913.092 in *w*PauO11, was then manually
714 masked. The Illumina reads used to assemble the draft *w*Pau genomes were then mapped
715 to the masked *w*PauO11 genome and the IS4 element consensus sequence using BWA-
716 mem v0.7.17-r1198-dirty. Repeated areas were identified from the annotation (see above)

717 and non-repeated regions of the *w*PauO11 genome were identified using “bedtools
718 complement” and the repeated regions as input. Coverage over the IS4 element consensus
719 sequence and the unmasked Undecim cluster copy was calculated with “samtools coverage”
720 and the coverage of the non-repeated regions of the genome was calculated by getting the
721 average per-base coverage in these regions by using “samtools depth”. The copy numbers
722 of the IS4 element and the Undecim cluster were calculated by dividing the average
723 coverage over these regions by the average coverage over the non-repeated regions of the
724 genome.

725 SNP phylogeny of *w*Pau genomes

726 To get a more resolved relationship between the *w*Pau strain variants, we called SNPs for all
727 *w*Pau strain variants from the Illumina reads mapped to the masked *w*PauO11 genome (see
728 above). Reads with a MQ<1 were filtered out and bcftools was used to call SNPs using the
729 command “bcftools mpileup --threads 6 -Ou -d 1000 -f $genome -b bam.file.list.txt | bcftools
730 call --threads 6 -cv --ploidy 1 -Ob -o variants_cflag.bcf”. To filter low-quality variants, the
731 command “bcftools view --types snps -i ‘QUAL>30 & INFO/DP>10 & MQ>30’” was used.
732 The variants located in non-repeated areas of the *w*PauO11 genome were extracted using
733 “bcftools view” with the coordinates identified as described above and “bcftools consensus”
734 was used to create pseudo-chromosomes for each strain. The sequence for each pseudo-
735 chromosome was put into a multi-fasta file and used as input for IQ-TREE v2.2.0 using the
736 same parameters as in the *Wolbachia* strain phylogeny. The phylogeny was visualized using
737 FigTree v1.4.4.

738 IS4 element insertion dynamics

739 To detect differences in IS4 insertions between the *w*Pau genomes, we identified improperly
740 mapped read pairs in the Illumina reads that were mapped to the masked *w*PauO11 genome
741 (see above). If an IS4 element is present in the genomes with mapped reads, the insert size
742 of paired reads in this region is 0, as one read of the pair would map on the masked
743 *Wolbachia* genome and the other read would map on the consensus IS4 sequence. To get
744 such improperly paired reads, we filtered with samtools view using parameters -F2, MQ>0
745 and -e “tlen==0”. Reads that mapped within 750 bases upstream of the start or 750 bases
746 downstream of the end of an IS4 element in wPauO11 (as specified in a provided bed file)
747 are from an IS4 element shared with *w*PauO11. Reads mapping at other places are
748 associated with IS4 insertions not present in *w*PauO11, and they were filtered by using
749 samtools view with the -U -L flag and the same bed file as above. If an IS4 element is
750 present in *w*PauO11 but absent in the genome with mapped reads, the reads from a pair are
751 further apart than expected, specifically the insert size would appear to be larger than the ca.

752 1500 bases of the IS4 element, and both reads of the pair should map on the masked
753 *Wolbachia* genome. To get such improperly paired reads, reads mapping 750 bases
754 upstream and 750 bases downstream of IS4 elements (as above) and that had an insert size
755 larger than 1500 bases but shorter than 5000 bases were filtered using samtools view -e
756 "(tlen>1500 && tlen<5000) || (tlen>-5000 && tlen < -1500)". The filtered read alignments of
757 improper pairs were visualized in IGV (v2.16.0) to identify unique IS4 element insertions
758 across the different *w*Pau strain variants. The *w*Pau whole genome SNP phylogeny was
759 then used to infer where and when IS4 copies were inserted in the *w*Pau strain variants.
760 Because complete deletions of IS4 elements occur rarely, differences were explained by
761 insertions, rather than deletions.
762 Additionally, we inferred the number of IS4 element copies present at each node of
763 the tree using complete *Wolbachia* genomes. We identified the shared IS4 elements across
764 the *w*Au-like strains, *w*Mel and *w*Cin2USA1 by blasting (blastn v2.7.1+) genomes to each
765 other and visualizing the regions flanking IS4 elements, identified by ISEscan, with Artemis
766 comparison tool (ACT) (Carver et al. 2005). An IS4 element was considered absent from a
767 genome if a region had the same gene order as an IS4 element-containing region of another
768 genome but didn't contain an IS4 element or if an IS4 element-containing region in other
769 genomes shared no synteny with the target genome.

770 Gene content variation between genomes *w*Au-like *Wolbachia*

771 To compare gene content across the different *Wolbachia* genomes, we used the gene count
772 output of Orthofinder to produce a hierarchical clustering of the genomes based on gene
773 presence/absence. First, we excluded all single-copy genes and then we assigned the count
774 of multi-copy genes to 1 so that we cluster based on presence and not copy-number. Then,
775 we used the "agnes" function for agglomerative clustering with the "average" method from
776 the cluster v2.1.4 package in R (Maechler et al. 2021).
777 Additionally, the variability in gene content between the wAu-like$_{ws}$ genomes was
778 investigated by identifying orthogroups where only one or several of the lineages, *w*Pau,
779 *w*Tro/*w*Wil and *w*Pro/*w*Spt, were present and one or several lineages were absent (Table 2,
780 Table S9 and Table S10). Unless otherwise stated, presence was counted if a gene was
781 present in the complete genome of the lineage (*w*PauO11, *w*WilP98 and *w*ProPE2) and at
782 least one draft genome of the same lineage.
783 All figures presenting genome and gene comparisons (Fig. 4A, Fig. 7, Fig. S7 and
784 Fig. S8) were made with genoPlotR (Guy et al. 2010), after using blastn to search each
785 genome against every other genome.

786 Prophage WO classification and gene module assignment

To investigate the prophage WO content in detail, we identified prophage WO regions in the complete *w*Au-like genomes, *w*PauO11, *w*ProPE2, *w*WilP98, *w*Inn, *w*Au, and in *w*Mel, *w*Cin2USA1. All proteins of prophage WO origin were re-clustered to orthogroups with OrthoFinder v2.5.4 with -I=1 and custom blastp (as above), which put most phage WO paralogues into single orthogroups. We manually validated the orthogroups based on the functional annotation and gene order in complete genomes (Table S13). The orthogroups containing proteins with the same functional annotation that were syntenic were manually merged. For orthogroup, we inferred a phylogeny using IQ-TREE v2.2.0 as described for single-copy genes. Genes were assigned to WOa and WOb based on whether they formed a clade with either WOMelA or WOMelB genes and/or were syntenic with WOMelA or WOMelB. For orthogroups where a WOMelB gene was present, we followed the classification of Bordenstein and Bordenstein 2022. Otherwise, we looked at what module the neighboring genes belonged to and assigned the orthogroup to that module. If an orthogroup had no functional annotation and could not be assigned to a module using the above method, it was assigned to the EAM module if it was flanked by genes in a structural module on one side and an EAM gene on the other side.

Acknowledgements

We thank Mercè Montoliu Nerín for assembling the *Drosophila paulistorum* O11 genome from which the *w*PauO11 genome contig was extracted and for helpful discussions throughout the project, Tom Martin for performing the DNA extraction for Nanopore sequencing of *D. paulistorum* O11, Lina Juzokaite for *Wolbachia*-enriched DNA extractions for PacBio sequencing, and Illumina sequencing. We also thank Athanasios Alexiou for assembling the first version of the *w*WilP98 genome. Sequencing was performed by the SNP&SEQ Technology Platform and Uppsala Genome Center (UGC) in Uppsala, Sweden. The facilities are part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. Some of the data handling was enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX. SNP&SEQ, UGC and Uppmax are supported by the Swedish Research Council and SNP&SEQ are supported by the Knut and Alice Wallenberg Foundation.

Author contribution

LK and WJM conceived and designed the study. KP and LK performed assemblies. KP and LK performed annotations. KP and LK performed comparative analyses. KP performed phylogenetic analyses. LK and KP wrote the paper with contributions from WJM. All authors read and approved the manuscript.

Data availability

The data underlying this study will be made available at publication.

823 # Figure legends

824 **Figure 1. Phylogeny of *Wolbachia* genomes with high similarity to *w*Au-like$_{ws}$ strains**.
825 The branches with *Wolbachia* strains associated with *Drosophila* species from the willistoni
826 and saltans groups, i.e., the *w*Au-like$_{ws}$ *Wolbachia* strains, are coloured. The clade referred
827 to as the *w*Au-like is indicated with an arrow. The tree was inferred from an alignment of 620
828 single-copy orthologs with IQ-TREE. Support values from 1000 ultrafast bootstrap replicates
829 are displayed on the nodes of the tree. Only bootstrap values greater than 60 are shown.

830 **Figure 2. Phylogenetic comparison between *w*Au-like *Wolbachia* and *Drosophila* of**
831 **the willistoni and saltans groups.** A) Maximum Likelihood Phylogenies of *w*Au-like
832 *Wolbachia* on the left and 692 nuclear-encoded proteins of the host on the right (from Baião
833 et al. 2023). The *w*Au-like *Wolbachia* phylogeny is a subset of the tree in Figure 1. B)
834 Maximum Likelihood Phylogenies of *w*Au-like *Wolbachia* on the left (same as in A) and the
835 whole mitochondrial genome of the neotropical *Drosophila* host species on the right (from
836 Baiaõ et al. 2023).

837 **Figure 3. Repeats and IS-elements in complete genomes of *w*Au-like strains**. A)
838 Percentage of repeats in each genome. B) Percentage of each IS family in the genomes, as
839 identified by ISEscan.

840 **Figure 4. Genomic rearrangements in *w*Au-like strains.** A) Whole genome alignments of
841 complete *Wolbachia w*Au-like genomes and outgroup. For each genome, genes are
842 visualized with blue and pseudogenes with green rectangles on either the forward or reverse
843 strand (above and below the midline respectively). Similarities between genomes are
844 visualized as red and blue ribbons, with red representing forward and blue reverse hits. B)
845 Gene order phylogeny of the same genomes as in A. Branch lengths written on each branch
846 represent the estimated number of genomic rearrangements.

847 **Figure 5. Gene content clustering of *Wolbachia w*Au-like genomes.** The cladogram on
848 the left shows how the *Wolbachia* strains cluster based on the presence and absence of
849 orthogroups in each genome, which is depicted in the plot to the right.

850 **Figure 6. Prophage WO genes in complete *w*Au-like genomes and outgroup.** Number
851 of genes per module for the two prophage WO variants identified in *w*Au-like and related
852 genomes. WOa genes were found in the same clade as genes from WO-A in *w*Mel and
853 WOb genes were found in the same clade as genes from WO-B in *w*Mel when inferring
854 single gene phylogenies.

855 **Figure 7. Duplication of the phage WO-associated genes in the Undecim cluster in**
856 ***w*PauO11.** Similarities between genomes are visualized with grey bands and the duplication
857 is marked in yellow. Genes in orange are part of the Undecim cluster. Mobile elements are
858 dark grey and pseudogenes green, with the border colour indicating Undecim cluster or
859 mobile element.

Tables

Table 1. Genomic features in complete *w*Au-like and outgroup genomes.

	***w*PauO11**[1]	***w*WilP98**[1]	***w*ProPE2**[1]	***w*Au**[2]	***w*Inn**[3]	***w*Mel**[4]	***w*Cin2**[5]
Genome size (bp)	1,365,198	1,268,527	1,339,466	1,268,455	1,290,587	1,267,782	1,538,351
GC (%)	34.8	35.2	35.2	35.2	35.3	35.2	35.1
CDSs	1000	1016	1033	996	1073	1011	1473
% Coding	73.8	77.5	76.3	75.7	70.1	76.0	86.2
Pseudogenes	177	123	130	122	227	111	216
Repeats (%)	22.9	11.4	15.8	11.0	16.8	9.8	13.6
Phage WO (%)	8.8	10.8	12.6	10.6	9.0	9.5	24.3
IS elements (%)	15.3	7.5	9.4	7.6	9.9	7.0	7.6

[1]*Wolbachia* genomes from Neotropical *Drosophila* sequenced in this study, [2] Baião GC et al., 2021, [3] Hill T et al., 2022, [4] Wu M et al., 2004, [5] Wolfe TM et al., 2021.

Table 2. Gene content variation between *w*Au-like$_{ws}$ *Wolbachia*. The numbers represent the presence or absence of an orthogroup in one lineage compared to the other two lineages in the table.

	***w*Pau**	***w*Wil/*w*Tro**	***w*Pro/*w*Spt**
Absent (total)	**66**	**18**	**12**
Phage WO	39	6	-
Ankyrin	2	1	2
Hypothetical	9	3	3
Dozen island	-	4	-
Present (total)	**6**	**11**	**16**
Phage WO	-	5	6
Ankyrin	1	-	1
Hypothetical	4	6	2
Dozen island	-	-	6

References

Baião GC, Janice J, Galinou M, Klasson L. 2021. Comparative Genomics Reveals Factors Associated with Phenotypic Expression of Wolbachia. *Genome Biology and Evolution* 13:evab111.

Baião GC, Schneider DI, Miller WJ, Klasson L. 2019. The effect of Wolbachia on gene expression in Drosophila paulistorum and its implications for symbiont-induced host speciation. *BMC Genomics* 20:465.

Baião GC, Schneider DI, Miller WJ, Klasson L. 2023. Multiple introgressions shape mitochondrial evolutionary history in Drosophila paulistorum and the Drosophila willistoni group. *Molecular Phylogenetics and Evolution* 180:107683.

Bakovic V, Schebeck M, Stauffer C, Schuler H. 2020. Wolbachia-Mitochondrial DNA Associations in Transitional Populations of Rhagoletis cerasi. *Insects* 11:675.

Baldo L, Dunning Hotopp JC, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, Hayashi C, Maiden MCJ, Tettelin H, Werren JH. 2006. Multilocus Sequence Typing System for the Endosymbiont Wolbachia pipientis. *Applied and Environmental Microbiology* 72:7098–7110.

Ballard JWO. 2004. Sequential Evolution of a Symbiont Inferred From the Host: Wolbachia and Drosophila simulans. *Molecular Biology and Evolution* 21:428–442.

Balvín O, Roth S, Talbot B, Reinhardt K. 2018. Co-speciation in bedbug Wolbachia parallel the pattern in nematode hosts. *Sci Rep* 8:8797.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477.

Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer ELL. 2002. The Pfam protein families database. *Nucleic Acids Res* 30:276–280.

Beckmann JF, Ronau JA, Hochstrasser M. 2017. A Wolbachia Deubiquitylating Enzyme Induces Cytoplasmic Incompatibility. *Nat Microbiol* 2:17007.

Blasco-Costa I, Hayward A, Poulin R, Balbuena JA. 2021. Next-generation cophylogeny: unravelling eco-evolutionary processes. Trends in Ecology & Evolution 36:907–918.

Bobay L-M and Ochman H (2017) The Evolution of Bacterial Genome Architecture. Front. Genet. 8:72. doi: 10.3389/fgene.2017.00072

Bordenstein Sarah R., Bordenstein Seth R. 2016. Eukaryotic association module in phage WO genomes from Wolbachia. *Nat Commun* 7:13155.

895 Bordenstein Sarah R., Bordenstein Seth R. 2022. Widespread phages of
896 endosymbionts: Phage WO genomics and the proposed taxonomic classification of
897 Symbioviridae.Matic I, editor. *PLoS Genet* 18:e1010227.

898 Bourtzis K, Nirgianaki A, Markakis G, Savakis C. 1996. Wolbachia infection and
899 cytoplasmic incompatibility in Drosophila species. *Genetics* 144:1063–1073.

900 Bruen TC, Philippe H, Bryant D. 2006. A Simple and Robust Statistical Test for
901 Detecting the Presence of Recombination. *Genetics* 172:2665–2681.

902 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.
903 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

904 Cariou M, Duret L, Charlat S. 2017. The global impact of Wolbachia on mitochondrial
905 diversity and evolution. *Journal of Evolutionary Biology* 30:2204–2210.

906 Charlat S, Duplouy A, Hornett EA, Dyson EA, Davies N, Roderick GK, Wedell N,
907 Hurst GD. 2009. The joint evolutionary histories of Wolbachia and mitochondria in
908 Hypolimnas bolina. *BMC Evolutionary Biology* 9:64.

909 Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, Barrell BG, Parkhill J.
910 2005. ACT: the Artemis Comparison Tool. Bioinformatics 21:3422–3423.

911 Charleston MA. 2016. Cospeciation. In: Encyclopedia of Evolutionary Biology.
912 Elsevier. p. 381–386.

913 Chrostek E, Marialva MSP, Yamada R, O'Neill SL, Teixeira L. 2014. High Anti-Viral
914 Protection without Immune Upregulation after Interspecies Wolbachia Transfer.
915 *PLOS ONE* 9:e99025.

916 Chrostek E, Teixeira L. 2015. Mutualism Breakdown by Amplification of Wolbachia
917 Genes.Malik HS, editor. *PLoS Biol* 13:e1002065.

918 Conner WR, Blaxter ML, Anfora G, Ometto L, Rota-Stabelli O, Turelli M. 2017.
919 Genome comparisons indicate recent transfer of Ri-like Wolbachia between sister
920 species Drosophila suzukii and D. subpulchrella. Ecology and Evolution 7:9391–
921 9404.

922 Cooper BS, Vanderpool D, Conner WR, Matute DR, Turelli M. 2019. Wolbachia
923 Acquisition by Drosophila yakuba-Clade Hosts and Transfer of Incompatibility Loci
924 Between Distantly Related Wolbachia. Genetics 212:1399–1419.

925 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A,
926 Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and
927 BCFtools. *GigaScience* 10:giab008.

928 De Vienne DM, Giraud T, Shykoff JA. 2007. When can host shifts produce congruent
929 host and parasite phylogenies? A simulation approach. Journal of Evolutionary
930 Biology 20:1428–1438.

931 Drillon G, Carbone A, Fischer G. 2014. SynChro: A Fast and Easy Tool to
932 Reconstruct and Visualize Synteny Blocks along Eukaryotic Chromosomes. *PLOS*
933 *ONE* 9:e92621.

934 Drillon G, Champeimont R, Oteri F, Fischer G, Carbone A. 2020. Phylogenetic
935 Reconstruction Based on Synteny Block and Gene Adjacencies. *Mol Biol Evol*
936 37:2747–2762.

937 Duarte EH, Carvalho A, López-Madrigal S, Costa J, Teixeira L. 2021. Forward
938 genetics in Wolbachia: Regulation of Wolbachia proliferation by the amplification and
939 deletion of an addictive genomic island. *PLOS Genetics* 17:e1009612.

940 Dudzic JP, Curtis CI, Gowen BE, Perlman SJ. 2022. A highly divergent *Wolbachia*
941 with a tiny genome in an insect-parasitic tylenchid nematode. *Proc. R. Soc. B.*
942 289:20221518.

943 Ellegaard KM, Klasson L, Näslund K, Bourtzis K, Andersson SGE. 2013.
944 Comparative Genomics of Wolbachia and the Bacterial Species Concept.Matic I,
945 editor. *PLoS Genet* 9:e1003381.

946 Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for
947 comparative genomics. *Genome Biology* 20:238.

948 Fisher RM, Henry LM, Cornwallis CK, Kiers ET, West SA. 2017. The evolution of
949 host-symbiont dependence. *Nat Commun* 8:15973.

950 Frank AC, Amiri H, Andersson SGE. 2002. Genome deterioration: loss of repeated
951 sequences and accumulation of junk DNA. *Genetica*.

952 Fricke LC, Lindsey ARI. 2024. Identification of Parthenogenesis-Inducing Effector
953 Proteins in Wolbachia. Genome Biology and Evolution 16:evae036.

954 Gavotte L, Henri H, Stouthamer R, Charif D, Charlat S, Boulétreau M, Vavre F. 2007.
955 A Survey of the Bacteriophage WO in the Endosymbiotic Bacteria Wolbachia.
956 *Molecular Biology and Evolution* 24:427–435.

957 Gerth M, Bleidorn C. 2016. Comparative genomics provides a timeframe for
958 Wolbachia evolution and exposes a recent biotin synthesis operon transfer. *Nat*
959 *Microbiol* 2:16241.

960 Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence
961 finishing. *Genome Res* 8:195–202.

962 Gutzwiller F, Carmo CR, Miller DE, Rice DW, Newton ILG, Hawley RS, Teixeira L,
963 Bergman CM. 2015. Dynamics of Wolbachia pipientis Gene Expression Across the
964 Drosophila melanogaster Life Cycle. G3 (Bethesda) 5:2843–2856.

965 Guy L, Roat Kultima J, Andersson SGE. 2010. genoPlotR: comparative gene and
966 genome visualization in R. Bioinformatics 26:2334–2335.

967 Hill T, Unckless RL, Perlmutter JI. 2022. Positive Selection and Horizontal Gene
968 Transfer in the Genome of a Male-Killing Wolbachia. Molecular Biology and
969 Evolution 39:msab303.

970 Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T. 2010. Wolbachia as a
971 bacteriocyte-associated nutritional mutualist. Proceedings of the National Academy
972 of Sciences 107:769–774.

973 Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, Analysis, and
974 Visualization of Phylogenomic Data. Molecular Biology and Evolution 33:1635–1638.

975 Ilinsky Y. 2013. Coevolution of Drosophila melanogaster mtDNA and Wolbachia
976 Genotypes. *PLOS ONE* 8:e54373.

977 Kampfraath AA, Klasson L, Anvar SY, Vossen RHAM, Roelofs D, Kraaijeveld K,
978 Ellers J. 2019. Genome expansion of an obligate parthenogenesis-associated
979 Wolbachia poses an exception to the symbiont reduction model. *BMC Genomics*
980 20:106.

981 Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence
982 alignment program. *Briefings in Bioinformatics* 9:286–298.

983 Kaur R, Shropshire JD, Cross KL, Leigh B, Mansueto AJ, Stewart V, Bordenstein
984 Sarah R., Bordenstein Seth R. 2021. Living in the endosymbiotic world of Wolbachia:
985 A centennial review. *Cell Host & Microbe* 29:879–893.

986 Kim BY, Wang JR, Miller DE, Barmina O, Delaney E, Thompson A, Comeault AA,
987 Peede D, D'Agostino ER, Pelaez J, et al. 2021. Highly contiguous assemblies of 101
988 drosophilid genomes.Coop G, Wittkopp PJ, Sackton TB, editors. eLife 10:e66405.

989 Lefoulon E, Bain O, Makepeace BL, d'Haese C, Uni S, Martin C, Gavotte L. 2016.
990 Breakdown of coevolution between symbiotic bacteria Wolbachia and their filarial
991 hosts. PeerJ 4:e1840.

992 Lefoulon E, Clark T, Guerrero R, Cañizales I, Cardenas-Callirgos JM, Junker K,
993 Vallarino-Lhermitte N, Makepeace BL, Darby AC, Foster JM, et al. 2020. Diminutive,
994 degraded but dissimilar: Wolbachia genomes from filarial nematodes do not conform
995 to a single paradigm. *Microb Genom* 6:mgen000487.

996 LePage DP, Metcalf JA, Bordenstein Sarah R., On J, Perlmutter JI, Shropshire JD,
997 Layton EM, Funkhouser-Jones LJ, Beckmann JF, Bordenstein Seth R. 2017.

998 Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic
999 incompatibility. *Nature* 543:243–247.

1000 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
1001 transform. *Bioinformatics* 25:1754–1760.

1002 Lynch M, Ali F, Lin T, Wang Y, Ni J, Long H. 2023. The divergence of mutation rates
1003 and spectra across the Tree of Life. EMBO reports 24:e57561.

1004 Lynch M, Conery JS. 2003. The origins of genome complexity. Science 302:1401–
1005 1404.

1006 Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. 2021. cluster: Cluster
1007 Analysis Basics and Extensions. Available from: https://CRAN.R-
1008 project.org/package=cluster

1009 Manzano-Marín A, D'acier AC, Clamens A-L, Orvain C, Cruaud C, Barbe V,
1010 Jousselin E. 2018. A Freeloader? The Highly Eroded Yet Large Genome of the
1011 Serratia symbiotica Symbiont of Cinara strobi. Genome Biology and Evolution
1012 10:2178.

1013 Manzano-Marín A, Latorre A. 2014. Settling Down: The Genome of Serratia
1014 symbiotica from the Aphid Cinara tujafilina Zooms in on the Process of
1015 Accommodation to a Cooperative Intracellular Life. Genome Biology and Evolution
1016 6:1683–1698.

1017 Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018.
1018 MUMmer4: A fast and versatile genome alignment system. *PLOS Computational*
1019 *Biology* 14:e1005944.

1020 Martinez J, Ant TH, Murdochy SM, Tong L, Filipe A da S, Sinkins SP. 2022. Genome
1021 sequencing and comparative analysis of Wolbachia strain wAlbA reveals Wolbachia-
1022 associated plasmids are common. *PLOS Genetics* 18:e1010406.

1023 Martinez J, Klasson L, Welch JJ, Jiggins FM. 2021. Life and Death of Selfish Genes:
1024 Comparative Genomics Reveals the Dynamic Evolution of Cytoplasmic
1025 Incompatibility. *Molecular Biology and Evolution* 38:2–15.

1026 Martinez J, Ok S, Smith S, Snoeck K, Day JP, Jiggins FM. 2015. Should Symbionts
1027 Be Nice or Selfish? Antiviral Effects of Wolbachia Are Costly but Reproductive
1028 Parasitism Is Not.McGraw EA, editor. *PLoS Pathog* 11:e1005021.

1029 McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria.
1030 *Nat Rev Microbiol* 10:13–26.

1031 Miller WJ, Ehrman L, Schneider D. 2010. Infectious Speciation Revisited: Impact of
1032 Symbiont-Depletion on Female Fitness and Mating Behavior of Drosophila
1033 paulistorum.Parrish C, editor. *PLoS Pathog* 6:e1001214.

1034 Miller WJ, Riegler M. 2006. Evolutionary Dynamics of *w* Au-Like *Wolbachia* Variants
1035 in Neotropical *Drosophila* spp. *Appl Environ Microbiol* 72:826–835.

1036 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A,
1037 Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
1038 Inference in the Genomic Era.Teeling E, editor. *Molecular Biology and Evolution*
1039 37:1530–1534.

1040 Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria.
1041 *Current Opinion in Genetics & Development* 14:627–633.

1042 Müller MJ, von Mühlen C, Valiati VH, Valente VL da S. 2012. Wolbachia pipientis is
1043 associated with different mitochondrial haplotypes in natural populations of
1044 Drosophila willistoni. *Journal of Invertebrate Pathology* 109:152–155.

1045 Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T. 2014.
1046 Evolutionary origin of insect–Wolbachia nutritional mutualism. Proceedings of the
1047 National Academy of Sciences 111:10257–10262.

1048 Perlmutter JI, Bordenstein Sarah R., Unckless RL, LePage DP, Metcalf JA, Hill T,
1049 Martinez J, Jiggins FM, Bordenstein Seth R. 2019. The phage gene wmk is a
1050 candidate for male killing by a bacterial endosymbiont. *PLOS Pathogens*
1051 15:e1007936.

1052 Powell JR, Sezzi E, Moriyama EN, Gleason JM, Caccone A. 2003. Analysis of a shift
1053 in codon usage in Drosophila. J Mol Evol 57 Suppl 1:S214-225.

1054 R Core Team. 2018. R: A Language and Environment for Statistical Computing.
1055 Vienna, Austria: R Foundation for Statistical Computing Available from:
1056 https://www.R-project.org/

1057 Rambaut, A., 2009. FigTree v1.4: Tree Figure Drawing Tool. Available at:
1058 http://tree.bio. ed.ac.uk/software/figtree/

1059 Raychoudhury R, Baldo L, Oliveira DCSG, Werren JH. 2008. Modes of Acquisition of
1060 *Wolbachia*: Horizontal Transfer, Hybrid Introgression, and Codivergence in the
1061 *Nasonia* Species Complex. *Evolution* 63:165–183.

1062 Renoz F, Arai H, Pons I. 2024. The genus Sodalis as a resource for understanding
1063 the multifaceted evolution of bacterial symbiosis in insects. Symbiosis 92:187–208.

1064 Renoz F, Foray V, Ambroise J, Baa-Puyoulet P, Bearzatto B, Mendez GL,
1065 Grigorescu AS, Mahillon J, Mardulyn P, Gala J-L, et al. 2021. At the Gate of
1066 Mutualism: Identification of Genomic Traits Predisposing to Insect-Bacterial
1067 Symbiosis in Pathogenic Strains of the Aphid Symbiont Serratia symbiotica. Front.
1068 Cell. Infect. Microbiol. 11.

1069 Russell SL, Chappell L, Sullivan W. 2019. A symbiont's guide to the germline. In:
1070 Current Topics in Developmental Biology. Vol. 135. Elsevier. p. 315–351.

1071 Ryan JA, Ulrich JM. 2022. quantmod: Quantitative Financial Modelling Framework.
1072 Available from: https://CRAN.R-project.org/package=quantmod

1073 Schneider DI, Ehrman L, Engl T, Kaltenpoth M, Hua-Van A, Le Rouzic A, Miller WJ.
1074 2019. Symbiont-Driven Male Mating Success in the Neotropical Drosophila
1075 paulistorum Superspecies. *Behav Genet* 49:83–98.

1076 Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*
1077 30:2068–2069.

1078 Sheeley SL, McAllister BF. 2009. Mobile male-killer: similar Wolbachia strains kill
1079 males of divergent Drosophila hosts. Heredity 102:286–292.

1080 Shen X-X, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic
1081 studies can be driven by a handful of genes. Nat Ecol Evol 1:1–10.

1082 Shropshire JD, Leigh B, Bordenstein SR. 2020. Symbiont-mediated cytoplasmic
1083 incompatibility: What have we learned in 50 years? *eLife* 9:e61989.

1084 Small ST, Tisch DJ, Zimmerman PA. 2014. Molecular epidemiology, phylogeny and
1085 evolution of the filarial nematode Wuchereria bancrofti. *Infect Genet Evol* 0:33–43.

1086 Syberg-Olsen MJ, Garber AI, Keeling PJ, McCutcheon JP, Husnik F. 2022.
1087 Pseudofinder: Detection of Pseudogenes in Prokaryotic Genomes. *Molecular Biology*
1088 *and Evolution* 39:msac153.

1089 Turelli M, Hoffmann AA. 1995. Cytoplasmic incompatibility in Drosophila simulans:
1090 dynamics and parameter estimates from natural populations. Genetics 140:1319–
1091 1338.

1092 Turelli M, Cooper BS, Richardson KM, Ginsberg PS, Peckenpaugh B, Antelope CX,
1093 Kim KJ, May MR, Abrieux A, Wilson DA, et al. 2018. Rapid Global Spread of wRi-like
1094 Wolbachia across Multiple Drosophila. Current Biology 28:963-971.e8.

1095 Unckless RL, Boelio LM, Herren JK, Jaenike J. 2009. Wolbachia as populations
1096 within individual insects: causes and consequences of density variation in natural
1097 populations. *Proceedings of the Royal Society B: Biological Sciences* 276:2805–
1098 2811.

1099 Unckless RL, Jaenike J. 2012. Maintenance of a Male-Killing Wolbachia in
1100 Drosophila Innubila by Male-Killing Dependent and Male-Killing Independent
1101 Mechanisms. Evolution 66:678–689.

1102 Vancaester E, Blaxter M. 2023. Phylogenomic analysis of Wolbachia genomes from
1103 the Darwin Tree of Life biodiversity genomics project.Teixeira L, editor. *PLoS Biol*
1104 21:e3001972.

1105 Wallau GL, da Rosa MT, De Ré FC, Loreto ELS. 2016. Wolbachia from Drosophila
1106 incompta: just a hitchhiker shared by Drosophila in the New and Old World? Insect
1107 Molecular Biology 25:487–499.

1108 Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview
1109 Version 2—a multiple sequence alignment editor and analysis workbench.
1110 Bioinformatics 25:1189–1191.

1111 Wernegreen JJ. 2015. Endosymbiont evolution: predictions from theory and
1112 surprises from genomes: Endosymbiont genome evolution. *Ann. N.Y. Acad. Sci.*
1113 1360:16–35.

1114 Werren JH, Baldo L, Clark ME. 2008. Wolbachia: master manipulators of
1115 invertebrate biology. *Nat Rev Microbiol* 6:741–751.

1116 Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G,
1117 Hayes A, Henry L, Hester J, et al. 2019. Welcome to the tidyverse. *Journal of Open*
1118 *Source Software* 4:1686.

1119 Wolfe TM, Bruzzese DJ, Klasson L, Corretto E, Lečić S, Stauffer C, Feder JL,
1120 Schuler H. 2021. Comparative genome sequencing reveals insights into the
1121 dynamics of Wolbachia in native and invasive cherry fruit flies. *Molecular Ecology*
1122 30:6259–6272.

1123 Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA,
1124 Martin W, Esser C, Ahmadinejad N, et al. 2004. Phylogenomics of the Reproductive
1125 Parasite Wolbachia pipientis wMel: A Streamlined Genome Overrun by Mobile
1126 Genetic Elements. PLOS Biology 2:e69.

1127 Xie Z, Tang H. 2017. ISEScan: automated identification of insertion sequence
1128 elements in prokaryotic genomes. *Bioinformatics* 33:3340–3347.

1129 Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular
1130 Biology and Evolution 24:1586–1591.

1131 Zeng Y, Wiens JJ. 2021. Do mutualistic interactions last longer than antagonistic
1132 interactions? Proceedings of the Royal Society B: Biological Sciences
1133 288:20211457.

1134 Zhang Y-K, Ding X-L, Zhang K-J, Hong X-Y. 2013. Wolbachia Play an Important
1135 Role in Affecting mtDna Variation of Tetranychus truncatus (Trombidiformes:
1136 Tetranychidae). *Environmental Entomology* 42:1240–1245.

Figure 1

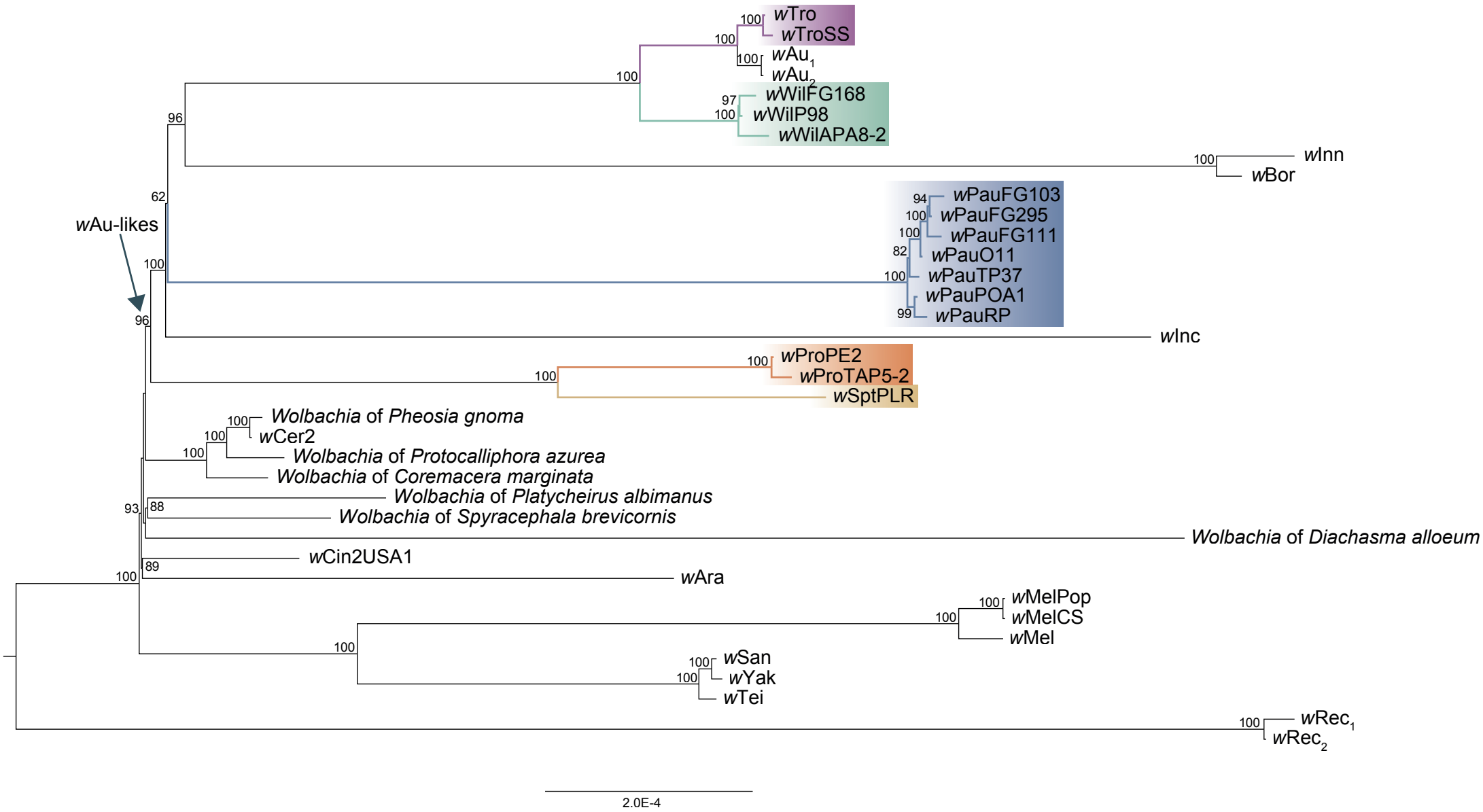
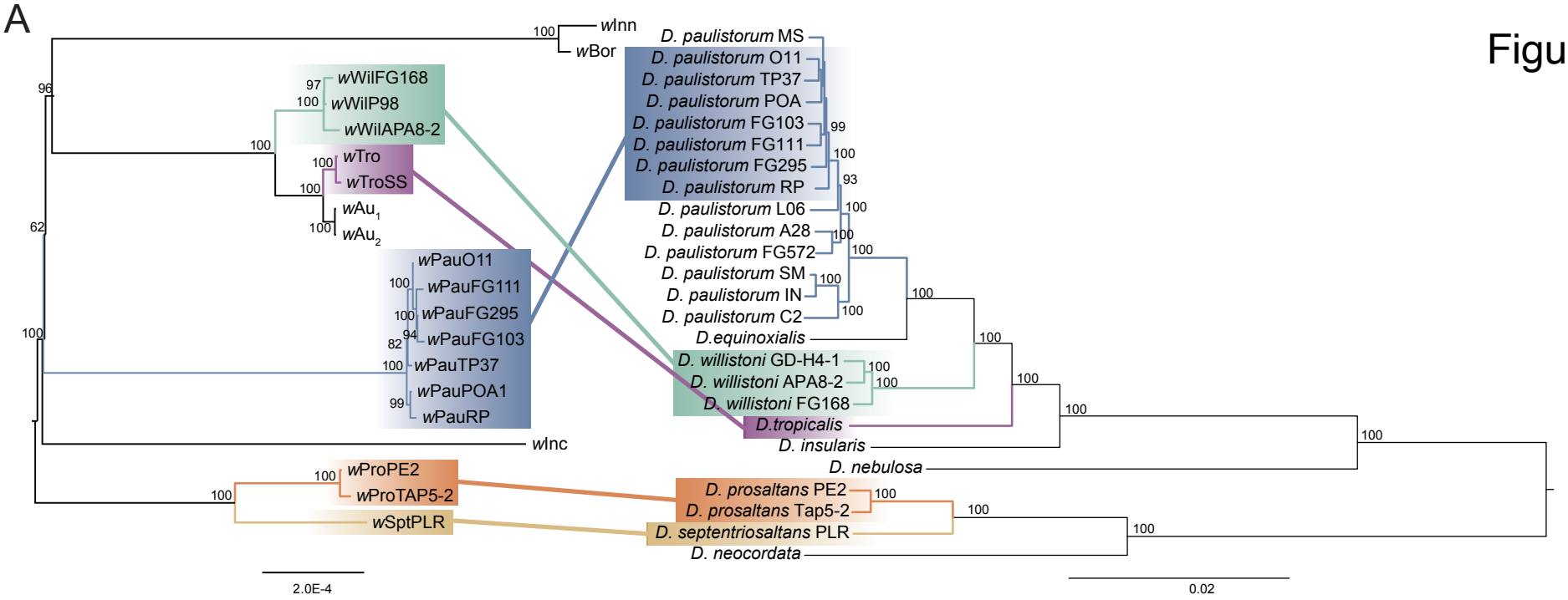

wTro
wTroSS
wAu1
wAu2
wWilFG168
wWilP98
wWilAPA8-2
wInn
wBor
wPauFG103
wPauFG295
wPauFG111
wPauO11
wPauTP37
wPauPOA1
wPauRP
wInc
wAu-likes
wProPE2
wProTAP5-2
wSptPLR
Wolbachia of Pheosia gnoma
wCer2
Wolbachia of Protocalliphora azurea
Wolbachia of Coremacera marginata
Wolbachia of Platycheirus albimanus
Wolbachia of Spyracephala brevicornis
Wolbachia of Diachasma alloeum
wCin2USA1
wAra
wMelPop
wMelCS
wMel
wSan
wYak
wTei
wRec1
wRec2
2.0E-4

Figure 2


A
wInn
wBor
wWilFG168
wWilP98
wWilAPA8-2
wTro
wTroSS
wAu1
wAu2
wPauO11
wPauFG111
wPauFG295
wPauFG103
wPauTP37
wPauPOA1
wPauRP
wInc
wProPE2
wProTAP5-2
wSptPLR
D. paulistorum MS
D. paulistorum O11
D. paulistorum TP37
D. paulistorum POA
D. paulistorum FG103
D. paulistorum FG111
D. paulistorum FG295
D. paulistorum RP
D. paulistorum L06
D. paulistorum A28
D. paulistorum FG572
D. paulistorum SM
D. paulistorum IN
D. paulistorum C2
D.equinoxialis
D. willistoni GD-H4-1
D. willistoni APA8-2
D. willistoni FG168
D.tropicalis
D. insularis
D. nebulosa
D. prosaltans PE2
D. prosaltans Tap5-2
D. septentriosaltans PLR
D. neocordata
2.0E-4
0.02

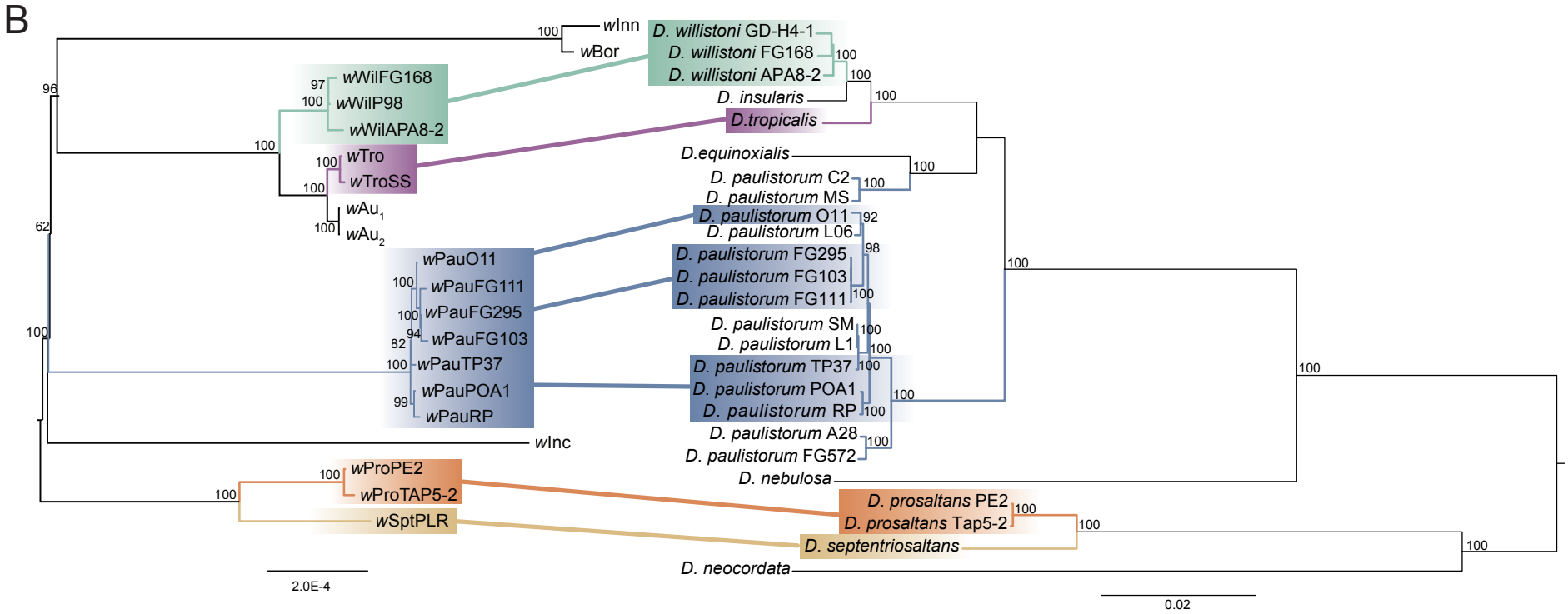

B
wInn
wBor
wWilFG168
wWilP98
wWilAPA8-2
wTro
wTroSS
wAu1
wAu2
wPauO11
wPauFG111
wPauFG295
wPauFG103
wPauTP37
wPauPOA1
wPauRP
wInc
wProPE2
wProTAP5-2
wSptPLR
D. willistoni GD-H4-1
D. willistoni FG168
D. willistoni APA8-2
D. insularis
D.tropicalis
D.equinoxialis
D. paulistorum C2
D. paulistorum MS
D. paulistorum O11
D. paulistorum L06
D. paulistorum FG295
D. paulistorum FG103
D. paulistorum FG111
D. paulistorum SM
D. paulistorum L1
D. paulistorum TP37
D. paulistorum POA1
D. paulistorum RP
D. paulistorum A28
D. paulistorum FG572
D. nebulosa
D. prosaltans PE2
D. prosaltans Tap5-2
D. septentriosaltans
D. neocordata
2.0E-4
0.02

Figure 3

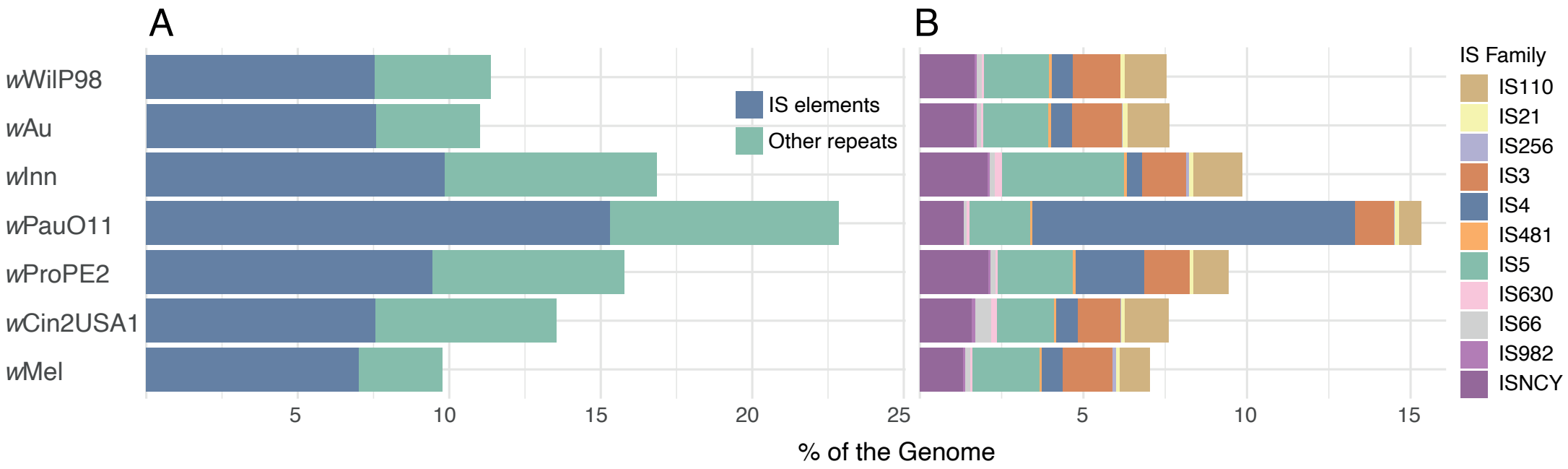

A
B
wWilP98
wAu
wInn
wPauO11
wProPE2
wCin2USA1
wMel
IS elements
Other repeats
5
10
15
20
25
5
10
15
% of the Genome
IS Family
IS110
IS21
IS256
IS3
IS4
IS481
IS5
IS630
IS66
IS982
ISNCY

Figure 4

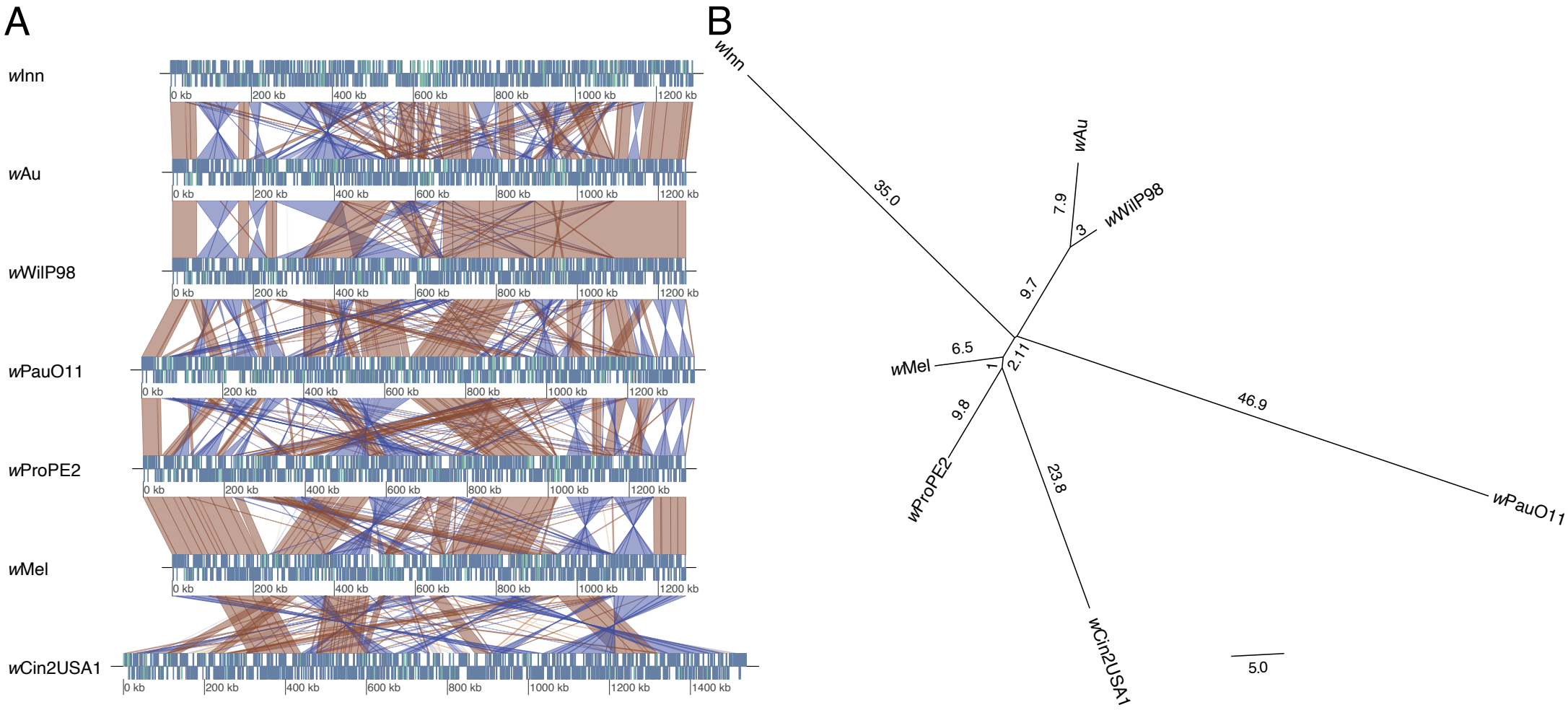

A
wInn
wAu
wWilP98
wPauO11
wProPE2
wMel
wCin2USA1
0 kb
200 kb
400 kb
600 kb
800 kb
1000 kb
1200 kb
1400 kb
B
wInn
wAu
wWilP98
wMel
wProPE2
wPauO11
wCin2USA1
35.0
7.9
3
9.7
6.5
1
2.11
9.8
46.9
23.8
5.0

Figure 5

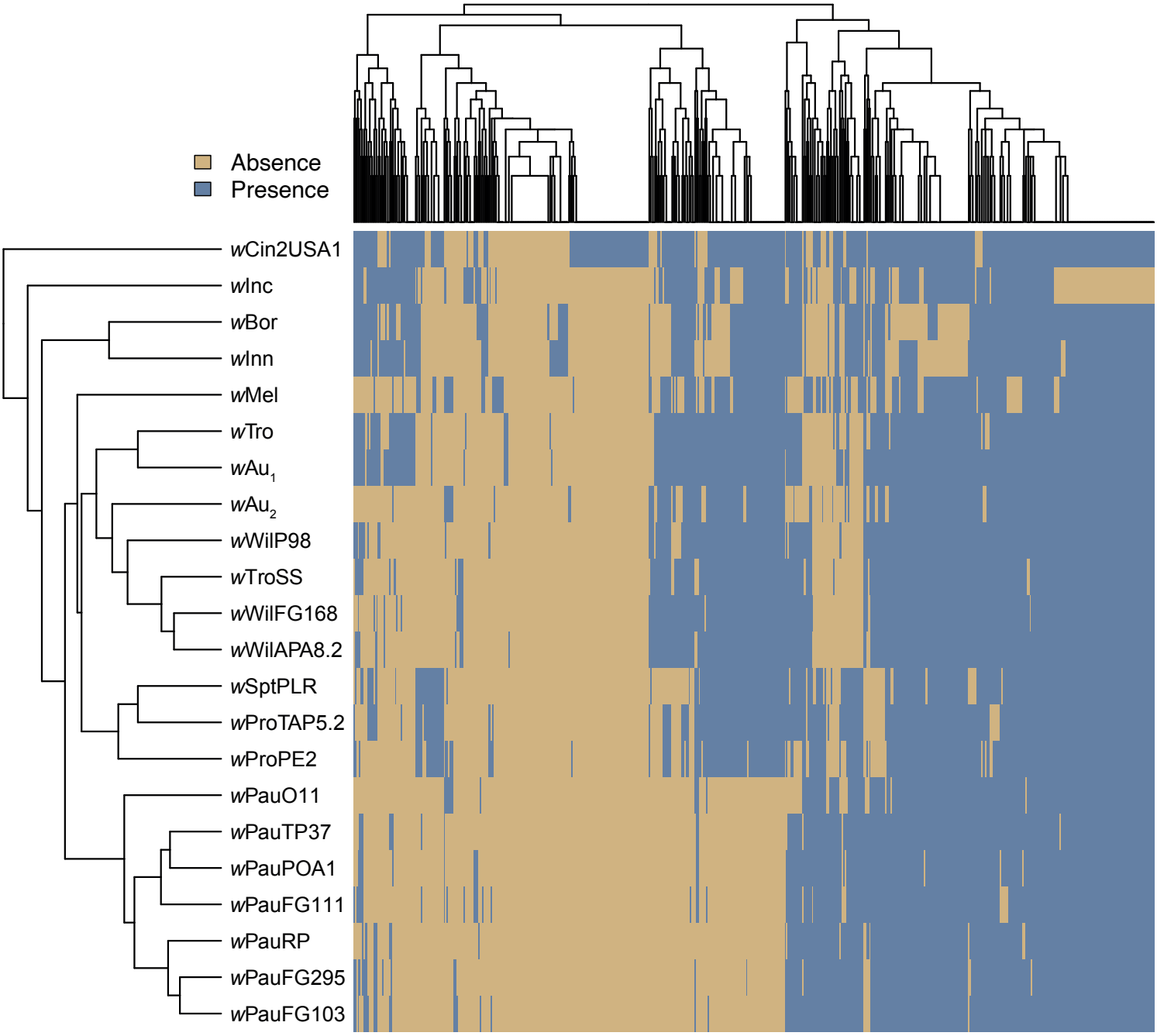

Absence
Presence
wCin2USA1
wInc
wBor
wInn
wMel
wTro
wAu1
wAu2
wWilP98
wTroSS
wWilFG168
wWilAPA8.2
wSptPLR
wProTAP5.2
wProPE2
wPauO11
wPauTP37
wPauPOA1
wPauFG111
wPauRP
wPauFG295
wPauFG103

Figure 6

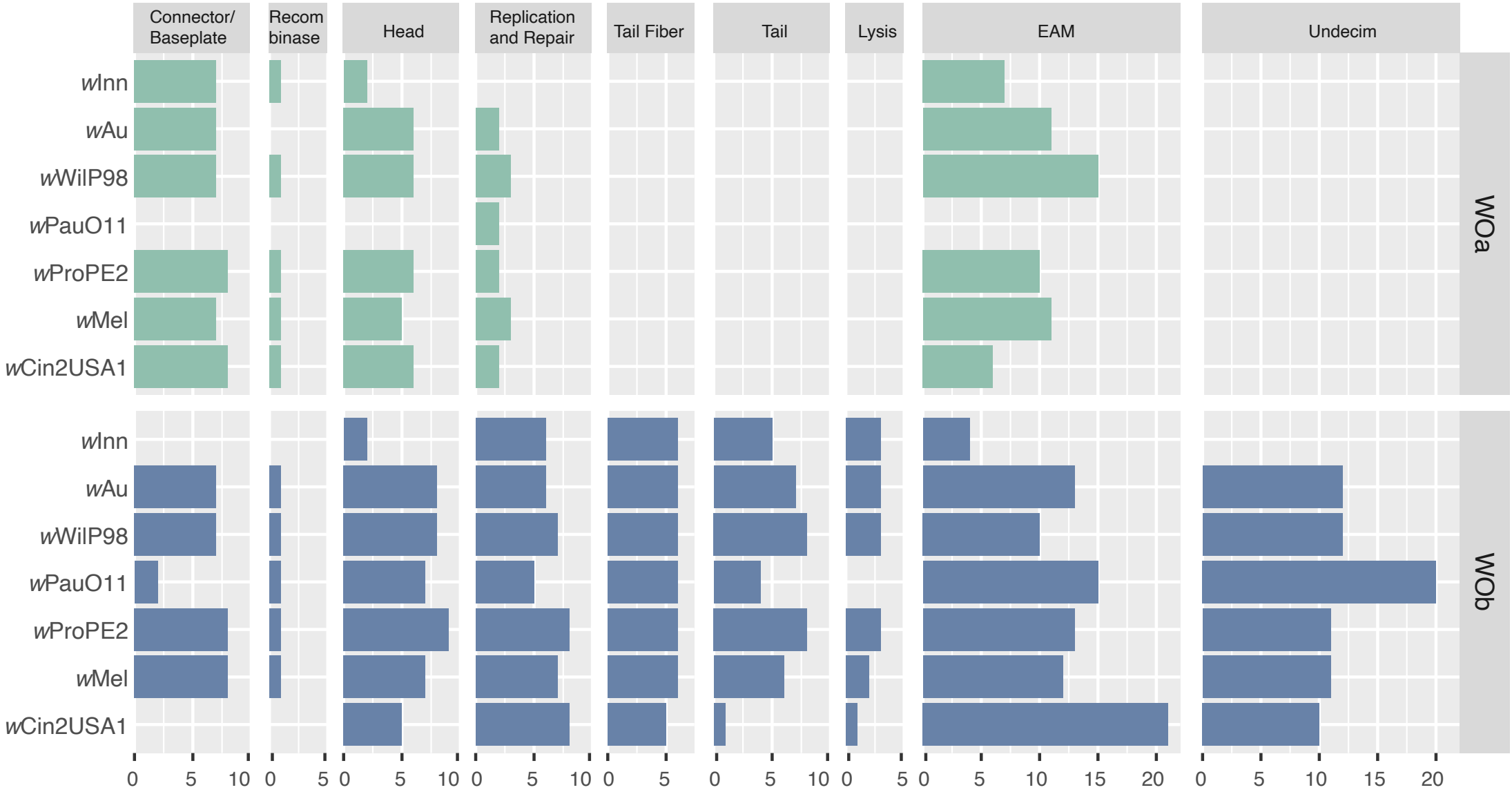
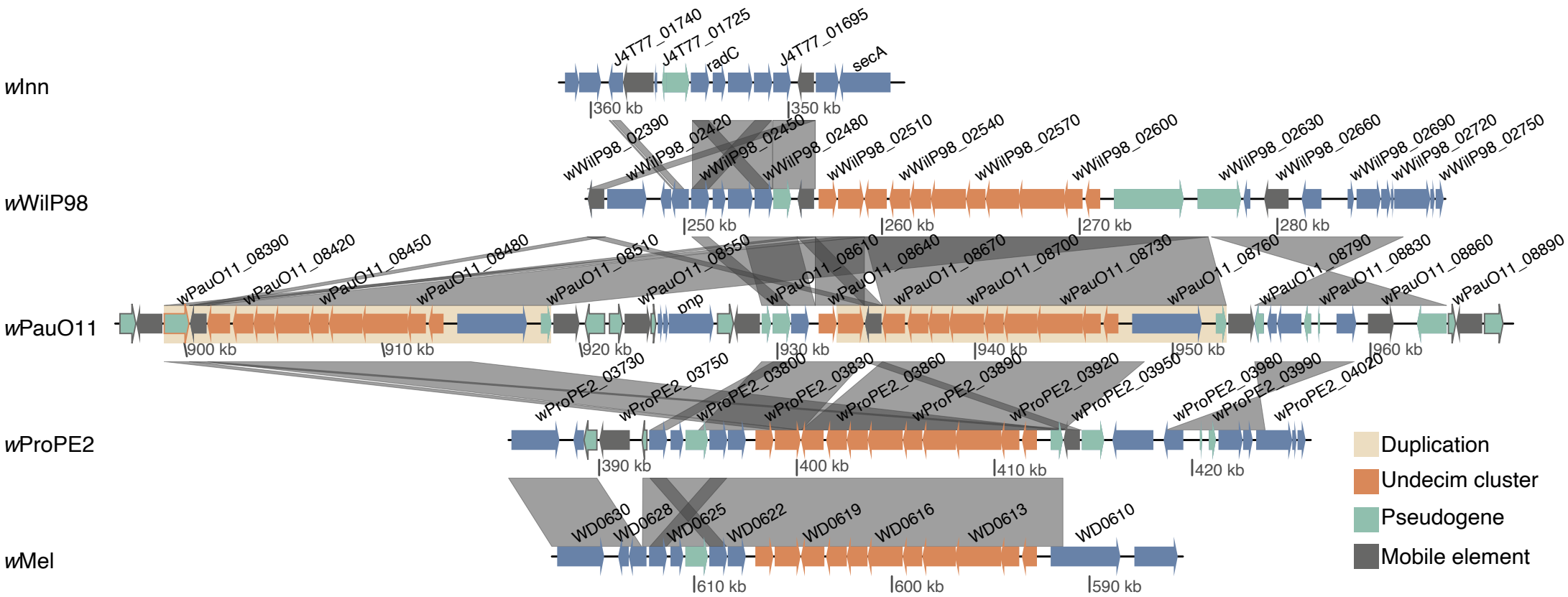

Connector/ Baseplate
Recom binase
Head
Replication and Repair
Tail Fiber
Tail
Lysis
EAM
Undecim
wInn
wAu
wWilP98
wPauO11
wProPE2
wMel
wCin2USA1
WOa
WOb
0
5
10
15
20

Figure 7


wInn
J4T77_01740
J4T77_01725
radC
J4T77_01695
secA
360 kb
350 kb
wWilP98
wWilP98_02390
wWilP98_02420
wWilP98_02450
wWilP98_02480
wWilP98_02510
wWilP98_02540
wWilP98_02570
wWilP98_02600
wWilP98_02630
wWilP98_02660
wWilP98_02690
wWilP98_02720
wWilP98_02750
250 kb
260 kb
270 kb
280 kb
wPauO11
wPauO11_08390
wPauO11_08420
wPauO11_08450
wPauO11_08480
wPauO11_08510
wPauO11_08550
pnp
wPauO11_08610
wPauO11_08640
wPauO11_08670
wPauO11_08700
wPauO11_08730
wPauO11_08760
wPauO11_08790
wPauO11_08830
wPauO11_08860
wPauO11_08890
900 kb
910 kb
920 kb
930 kb
940 kb
950 kb
960 kb
wProPE2
wProPE2_03730
wProPE2_03750
wProPE2_03800
wProPE2_03830
wProPE2_03860
wProPE2_03890
wProPE2_03920
wProPE2_03950
wProPE2_03980
wProPE2_03990
wProPE2_04020
390 kb
400 kb
410 kb
420 kb
wMel
WD0630
WD0628
WD0625
WD0622
WD0619
WD0616
WD0613
WD0610
610 kb
600 kb
590 kb
Duplication
Undecim cluster
Pseudogene
Mobile element