Pseudoscorpion Wolbachia symbionts: Diversity and Evidence for a New Supergroup S

Emilie Lefoulon
New England Biolabs Inc
Travis Clark
New England Biolabs Inc
Fanni Borveto
Universite de Montpellier
Marco Perriat-Sanguinet
Universite de Montpellier
Catherine Moulia
Universite de Montpellier
Barton Slatko (slatko@neb.com)
New England Biolabs
Laurent Gavotte
Universite de Montpellier

Research article

Keywords: Wolbachia, pseudoscorpion, symbiosis, target enrichment, genomics, biotin

Posted Date: January 29th, 2020

DOI: https://doi.org/10.21203/rs.2.22118/v1

Version of Record: A version of this preprint was published at BMC Microbiology on June 30th, 2020. See the published version at https://doi.org/10.1186/s12866-020-01863-y.

Abstract

Background

Wolbachia are the most widely spread endosymbiotic bacteria, present in a wide variety of insects and two families of nematodes, but so far, relatively little genomic data is available. The *Wolbachia* symbiont can be a parasite, as described for many arthropods, or an obligate mutualist, as in filarial nematodes. Although, the nature of these symbioses remains largely unknown, diverse *Wolbachia* genomic data will contribute to understanding their diverse symbiotic mechanisms. Results Our study focuses on *Wolbachia* infections in pseudoscorpion species collected in Montpellier (France) and indicates two distinct groups of *Wolbachia* : *Geogarypus minor* harbors *Wolbachia* (w Gmin) and *Chthonius ischnocheles* harbors *Wolbachia* (w Cisc), both related to supergroup H and Atemnus politus harbours *Wolbachia* (w Apol), forming a novel divergent clade with *Wolbachia* from the pseudoscorpion *Cordylochernes scorpioides* . This later clade forms a new supergroup S, most closely related to *Wolbachia* supergroups C and F. Our data suggest multiple symbiont acquisition events within the evolutionary history of pseudoscorpions. Using target enrichment by hybridization with *Wolbachia* -specific biotinylated probes to capture large fragments of *Wolbachia* DNA, we produced a draft genome of w Apol characterized by a moderate *Wolbachia* size (1,445,964bp) containing a moderate number of transposable elements and WO bacteriophage insertions (total of 77,522 bp). Conclusions Our analyses indicate that w Apol forms a diverget clade. Annotation highlights complete biochemical pathways which are incomplete in many sequenced *Wolbachia* genomes, such as vitamin B and the cytochrome bd ubiquinol oxidase pathway. Further, the biotin operon appears to have been horizontally transferred multiple times along the *Wolbachia* evolutionary history.

Background

Wolbachia are endosymbiotic alpha-proteobacteria infecting a broad range of arthropods and nematodes [1, 2]. The bacteria of this genus are considered to be the most widely spread symbionts in the animal world, perhaps infecting half of insect species [3-5]. *Wolbachia* are maternally inherited and can induce variable phenotypes in their hosts through mutualism or parasitism [6-9]. *Wolbachia* are genetically diverse, as are the interactions with their hosts [10, 11]. Currently, there is a general consensus to classify them in monophyletic lineage groups ("supergroups" A to R). *Wolbachia* belonging to supergroups C, D and J exclusively infect filarial nematodes (Onchocercidae) [12-14]. Supergroup L exclusively contains plant parasitic nematodes (Pratylenchidae) [15, 16]. Supergroup F *Wolbachia* is, so far, the only clade composed by some strains infecting arthropods and some infecting filarial nematodes [17, 18]. All other described supergroups exclusively infect arthropods [19-23]. The proposed supergroup G members [24] and R [25] are now considered to be part of supergroup B [26] and A [27] and are no longer considered valid as separate supergroups.

In the last few years, the number of published *Wolbachia* genomes has greatly increased. Currently 48 draft and 21 complete genomes of *Wolbachia* are available at the NCBI database. However, these data

are not a good representation of the entirety of *Wolbachia* diversity. Indeed, among the 21 complete genomes, 15 are symbionts of insects belonging to either supergroup A or B (including 7 strains from species of *Drosophila*), 2 are symbionts of insects belonging to other supergroups (F and E) and 4 are symbionts of nematodes (supergroups L, C or D). *Wolbachia* has been identified in other arthropods, such as isopods [28, 29] and arachnids [24]. Some genomic data for isopods are available, for example a draft genome of *w*Con, infecting *Cylisticus convexus* and a draft genome of *w*VulV infecting *Armadillidium vulgare* [30]. No genomics data is available for arachnid symbionts while several studies identified *Wolbachia* in spiders [24, 25, 31], mites [20] or scorpions [32].

In the current study, we focused on symbionts of pseudoscorpions. The presence of *Wolbachia* was reported for the first time in pseudoscorpions in 2005 in *Cordylochernes scorpioides* [33]. We examined the prevalence of *Wolbachia* infections in pseudoscorpion population samples from an area in Montpellier (France) where three different species of pseudoscorpions were collected and determined to be positive for *Wolbachia* infection. The molecular analyses phylogenetically identified two different *Wolbachia* groups, of which one strain is divergent from currently accepted *Wolbachia* supergroups. We used *Wolbachia* DNA enrichment capture [34, 35] to ensnare *Wolbachia* DNA for genomic sequencing from this *Wolbachia* strain, *wApo,* infecting *Atemnus politus*.

Results

Identification and Evolution of Pseudoscorpions

In all, 94 pseudoscorpion specimens were collected from an area of Montpellier (France). 60 specimens were morphologically identified as *Atemnus politus* (Simon, 1878), 24 specimens belong to the species *Geogarypus minor* (Koch, 1873) and 10 specimens belong to the species *Chthonius ischnocheles* (Hermann, 1804). The specimen identification is consistent with the analysis of the mitochondrial cytochrome oxidase I gene marker (COI) as the COI sequences of the specimens belonging to *Geogarypus minor* are identical to each other and to the sequence of *Geogarypus nigrimanus* (JN018180 specimen voucher MNHN-JAB62). The species *Geogarypus nigrimanus* has been recently synonymized with *Geogarypus* minor [36]. Regarding the produced COI sequences of the specimens belonging to *Chthonius ischnocheles* species, the specimens IV3-1, Q3-1, IV1-1 and Q4 are between 99.8% -100% identical and match the sequence of *Chthonius ischnocheles* available at the NCBI database (JN018172 specimen voucher MNHN-JAB62) but the specimens IV-J5 and IV-S1 form a paraphyletic group with the other *Chthonius ischnocheles* (Figure 1). However, the existence of cryptic species of *Chthonius ischnocheles* has been documented [37]. The produced COI sequences of the specimens identified as *Atemnus politus* present more variability: calculated pairwise similarities are ranging between 97.7-99.8%. The nblast analysis does not allow the identification of strongly similar sequences in databases, indeed no COI sequences from species related to the *Atemnus* genus have been deposited in the NCBI database. The most similar COI sequence in the database belongs to *Atemnidae* sp. JA-2011 voucher (JN018203), showing only 79.74% similarity. The COI-based phylogeny indicates that all the specimens of *Atemnus politus* form a clade which is a sister group of the clade composed of *Paratemnoides sumatranus* and

Oratemnus curtus species, both representatives of the *Atemnidae* family (Figure 1). The COI analysis shows all specimens of *Chthonius ischnocheles* form a sister group of other species belonging to *Chthonius* species (*C. tetrachelatus*, *C. dacnodes*) (Figure 1). To date, the only species of pseudoscorpion described as a *Wolbachia* host, *Cordylochernes scorpioides*, belongs to the Chernetinae and is sister group of representative of the Atemnidae family (including *Atemnus politus*) (Figure 1).

Atemnus politus mitochondria organization

We were also able to assemble the complete the mitochondrial genome sequence of *Atemnus politus* and compare its organization with two complete mitochondria previously sequenced: *Paratemnoides elongates* (JQ040543) and *Pseudogarypus banksi* (JQ040544) (Ovchinnikov and Masta 2012) (Additional File 1, Figure S1). The circular mitochondria encodes the typical set of 13 protein-coding genes and two ribosomal RNA subunits but only 16 tRNA genes (trnA, trnC, trnG, trnH, trnI, trnT are absent). The mitochondrial genome arrangement observed in *A. politus* is close to that of of *P. elongates*, with the exception that some tRNA genes are absent (trnN, trnT, trnW for *P. elongates*) and the *nad*6 gene is split in *P. elongates*.

Prevalence of *Wolbachia* infection and sequencing data

Wolbachia was detected by PCR amplification of *wsp* and *ftsZ* markers (see Methods) in the three studied pseudoscorpion species and in accordance with the generally used nomenclature of *Wolbachia* strains, we have named these new strains according to their hosts: *wApol* for bacteria infecting *Atemnus politus*, *wGmin* for bacteria infecting *Geogarypus* minor and *wCisc* for bacteria infecting *Chthonius ischnocheles*.

The *A. politus* specimens have the highest prevalence of *Wolbachia* infection with 43% positive samples (60 specimens examined). 17% of *G. minor* individuals appear to be infected (24 specimens examined), while only 10% of the *C. ischnocheles* (10 specimens examined) show PCR amplification of the *Wolbachia* markers. However, the density of infection in *G. minor* could be much lower than in the other two. In fact, nested PCR amplifications were necessary to obtain *Wolbachia* sequences for DNA analysis.

A draft genome was produced for *wApol Wolbachia* from *Atemnus politus* (specimen K5) using target capture enrichment [34, 35]. For Illumina sequencing, an enrichment of 45-fold was observed (0.4% of the reads mapped to the draft genome without enrichment against 18% with) and 50-fold enrichment was observed with PacBio sequencing (0.8% of the read mapped to the draft genome without enrichment against 40% with). The hydrid *de novo* assembly allowed production of a 373 contig draft genome, 1,445,964 bp long with an average G+C content of 35.6% (largest contig, 25,286 bp, N50=5,741 bp, average sequencing depth= 420X) (Table 1).

Table 1: Information for the wApol K5 **genome sequence.**

Summary of *de novo* assembly using canu processed in the current study: statistics using QUAST (Gurevich, Saveliev et al. 2013); annotation using RAST*(Aziz, Bartels et al. 2008)

pipeline or prokka**(Seemann 2014); assessment of the draft using BUSCO v3 (Seppey, Manni et al. 2019).

	wApol K5
Number of contigs	373
Size of the largest contig	25,286
Total length (bp)	1,445,964
Contigs >= 10,000bp	28
N50	5,741
L50	73
GC%	35.61
Number of Coding Sequences	1746*/1803**
Number of RNAs	39*/38**
Complete BUSCOs %	70.2%
Fragmented BUSCOs %	7.2%
Missing BUSCOs %	22.6%

Multi-locus phylogenies and phylogenomics

The current unrooted *Wolbachia* phylogeny, based on the concatenation of six genes (16S rDNA, *ftsZ*, *dnaA*, *gatB*, *fbpA* and *coxA*), identified two different groups of *Wolbachia* in our samples (Figure 2A, Figure 2B). The *Wolbachia* infecting *Chthonius ischnocheles* appears closely related to supergoup H *Wolbachia* infecting *Zootermopsis* termite species. Only two genes were sequenced (*coxA* and *ftsZ*) for the *Wolbachia* infecting *Geogarypus minor* and the analysis of these genes shows *w*Gmin is closely related to *w*Cisc (Figure 2B). The *Wolbachia* from *Atemnus politus* is divergent from known supergroups and is a sister group to supergroup C *Wolbachia* which contains the symbiont present in filarial nematodes (*Onchocerca* spp. and *Dirofilaria immitis*).

The only strain of *Wolbachia* known to infect pseudoscorpions, *Wolbachia* from *Cordylochernes scorpioides,* was not included with the first multi-locus phylogeny because only 2 of the 6 genes had been previously sequenced (*coxA* and *ftsZ).* Indeed, this *Wolbachia* has been studied using a set of genes not routinely sequenced for *Wolbachia* studies (*groEL*, *fabK*, *nuoG*, *NADH dehydrogenase I subunit F*, *aspS*, *gltA*, *coxA*, *ftsZ*, *wsp*, *orpB*, *nuoD*, isocitrate dehydrogenase gene, TPR domain-containing protein gene) [38]. We performed a multi-locus phylogeny based on these 13 genes, including the *Wolbachia* from *Cordylochernes scorpioides* and sequences isolated from available complete or draft genomes of *Wolbachia*, as well as the produced draft genome *wApol*. The observed phylogeny shows that *Wolbachia* from *Cordylochernes scorpioides* and the symbiont from *Atemnus politus wApol* form a clade together as a closely related sister group to supergroup C. The two pseudoscorpions harboured closely related *Wolbachia* strains, with 89.3% identity among these 13 genetic markers (Additinal file 2, Figure S2).

Two complete phylogenomic analyses were performed: one including 16 *Wolbachia* genomes representing 6 different supergroups and one including 33 *Wolbachia* genomes representing 7 different supergroups, also including additional draft genome sequences. 374 single-copy orthologues were identified among the 16 *Wolbachia* genomes, while 205 were identified among the 33 *Wolbachia* genomes. The Maximum Likelihood phylogeny based on the reduced matrix (16 genomes, 374 orthologues, 101,994 amino-acid sites) (Figure 3) and the one based on the larger matrix (37 *Wolbachia*, 192 orthologues, 45,930 amino-acid sites) (Figure 3) present *w*Apol group with any other studied *Wolbachia* suggests it evolved from an independent speciation event. Based upon the multi-locus phylogenies (Figure 3), *wApol* K5 is closely related to the supergroup C (symbionts within the filarial nematodes *Onchocerca* spp. and *Dirofilaria immitis*) and the supergroup F (including both filarial and arthropods symbionts).

The *wApol* draft genome and its annotation

Among 221 single-copy orthologue genes conserved among proteobacteria (BUSCOs database), 155 are present in the *wApol* draft genome, suggesting a 70.2% complete BUSCOs. As described in the Methods section, this percentage can be used to assess the level of completeness of genomes. By comparison, the *Wolbachia* from *Drosophila melanogaster, w*Mel, has a higher level of completion in the current analysis with 180 genes (81.4%) and *Wolbachia* from *Pratylenchus penetrans, w*Ppe, has a lower level with 161 genes (72.9%) (Additional file 3, Figure S3, Additional file 4, Table S1).

The Average Nucleotide Identity (ANI) calculation (see Methods) indicates that *wApol* is most similar to *wCle* with 87% identity (Figure 4). This is consistent with the phylogenetic proximity of the two strains. According to the analysis of available complete genomes, strains representative of the same supergroup share more than 90% identity; for example, 99% for *wOo* and *wOv* (from supergroup C), 94% for *wPip* and *wTpre* (from supergroup B) or 96% for *wMel* and *wCau* (from supergroup A). The *in-silico* genome-to-genome comparison might suggests that *w*Apol represents a different *Wolbachia* species. Indeed, the digital DNA-DNA hybridization (DDH)[39] higher than 70 indicates that the two strains might belong to the same species (see Methods). The current analysis indicates only *wOo* with *wOv* and *wMel* with *wCau* having a dDDH higher than 70, representing the same *Wolbachia* species, as previously suggested [40].

1746 coding Sequences (CDS) (373 with protein-encoding genes PEGs) were annotated from *wApol* using RAST while 1803 CDS were found using prokka. With respect to transposable elements, 52 insertion sequences (ISs) were detected (Figure 5, Additional file 5, Table S2) as were 55 mobile elements and 16 Group II intron-associated genes. These represent a low number of transposable elements compared to other *Wolbachia* infecting insects (with the exception of *wTpre*), but a high number compared to *Wolbachia* infecting filarial nematodes (Figure 5, Additional file 6, Table S3, Additional file 7, Table S4). The ISs were diverse, with 8 different IS families. The IS5 and IS630 families appear overrepresented (respectively n=19 and n=12). Most analyzed *Wolbachia* genomes contain the IS5 family; in particular, *w*Cle, *w*Pip, *w*Mel, *w*Nfla and *w*CauA contain a high level of IS5. *w*Apol contains a moderate number of mobile elements (n=55) compared to other *Wolbachia* infecting insects, but a high

number compare to *Wolbachia* infecting filarial nematodes (Figure 5, Additional file 7, Table S4). *w*Apol contains a high number of group II introns, more than the other studied *Wolbachia* genomes (with the exception of *w*stri and *w*VulC which present a higher number) (Figure 5).

Metabolic pathways

Functional annotation of genes by KAAS highlights differences among studied *Wolbachia* genomes (Figure 6). Of interest is the observation that some genes identified in *w*Apol are absent in most of the other studied *Wolbachia*, such as two cases in the vitamin B metabolism pathway involved in biotin and thiamine metabolism. With respect to biotin metabolism, only *w*Cle (supergroup F), *w*Nfla, wNleu (supergroup A) and *w*Lug, *w*VulC (close to the supergroup B) and *w*Apol have a complete or almost complete pathway (only *bioC* is not detected in *w*Apol). The phylogenies of the biotin genes (bioA, bioB, bioC, bioD and bioF) are not congruent with the evolutionary history of *Wolbachia* and appear to be derived from several acquisitions (Figure 9). As previously suggested [41], our data indicates that the biotin operon in *Wolbachia* appears to be closely related to other symbiotic bacteria, in particular *Cardinium* species. This suggests that this operon might be acquired by multiples lateral transfer event among the *Wolbachia* evolutionary history. A second example is the thiamine metabolism pathway, where the *tenA* gene was observed in *w*Apol, but not in *Wolbachia* infecting filarial nematodes (supergroups C and D) or *w*Tpre (supergroup B), but is present in *w*Mel, *w*CauA, wNfla, wNleu (all from supergroup A), *w*Pip (supergroup B), *w*Fol (supergroup E) and *w*Cle (supergroup F). Regarding the oxidative phosphorylation pathway, there are differences with cytochrome bd ubiquinol oxidase genes; the *cydA* and *cydB* genes were detected only in *w*Apol and in *w*Mel, *w*CauA, *w*Nleu and wNfla (supergroup A). Numerous pathways are present that are potentially involved in *Wolbachia*-induced host phenotypes, such as the effector type IV secretion system (T4SSs), considered to be an effector translocator (Pichon, Bouchon et al. 2009). T4SS was identified in *w*Apol as clustered in two loci, similar to other *Wolbachia* strains [42]: five tandem genes virB8, -B9, -B10, -B11 and virD4 homologs were identified in one contig (#25); three tandem genes virB3, -B4 and -B6 were identified in another contig (#16) (Additional file 8, Figure S4). Some pilus-associated proteins, such as VirB2 (major pilus subunit of type IV secretion complex) were identified in different contigs (#6, #42 and #127 contigs). Interestingly, these genes involved in host cell attachment have been described as lost in Rickettsiales genomes as well as *w*Mel and *w*Ri (Pichon, Bouchon et al. 2009)[43]. More recently, it has been highlighted that some *Wolbachia* genomes contain VirB2, such as some strains infecting *Drosophila* species [44], those infecting *Dactylopius* species [45] and *Wolbachia* from *Aedes albopictus* [46]. Our results indicate presence of VirB2 in all *Wolbachia* analyzed, with the exception of *w*Ppe, and only one copy is observed in the genomes of *w*Oo, *w*Ov and *w*Bm (Additional file 9, Table S5). The entire bacterial secretion system is conserved in *w*Apol (with the exception of the gene *secY* from the Sec-SRP system) (Figure 6). The SS type II gene (*gspD*) is found in *w*Apol draft but is not conserved in *Wolbachia* from filarial nematodes (*w*Oo, *w*Ov and *w*Bm). Similarly, the cell cycle pathway is conserved in *w*Apol, similar to other *Wolbachia* (with the exception of the *ftsQ* division cell gene) (Figure 6).

Presence of phage region in *wApol*

Numerous phage genes were found in the *w*Apol draft genome. 48 of the 373 contigs of *wApol* contain regions of strong similarity with WO bacteriophage sequences, representing a total of 77,522 bp (smallest read is 1,001 nts and the largest read is 1,642 nts) (Additional file 10, Table S6). 13 regions similar to WO bacteriophage are conserved, coding mainly ankyrin or tetratricopeptide repeat family proteins, hypothetical proteins or structural phage proteins (tail, portal phage, capsid) (Additional file 10, Table S6). Among these regions, 6 were confirmed as similar to prophage sequences using PHASTER [47]. A Maximum Likelihood analysis of these regions (Additional file 11, Figure S5; Additional file 12, Figure S6; Additional file 13, Figure S7), indicates that these contigs (listed in Additional file 10, Table S6) do not appear to share the same evolutionary history. Some appear more closely related to regions present in *w*Cle, some are more related to regions of *w*Fol, and others are more related to regions of diverse *Wolbachia* from supergroups A or B or even clades containing both supergroups A and B.

Some genes associated with WO bacteriophage have been described as being involved in reproductive parasitism of *Wolbachia*. This the case for *cifA* and *cifB* involved in the cytoplasmic incompatibility phenotype [48] or the *wmk* gene as a candidate gene for male-killing [49]. No orthologues of WD0631 (*cifA* gene of *wMel*) or WD0632 (*cifB* gene of *wMel*) were identified in the *wApol* draft genome. Of potential interest is that, while only a small length of sequence, the beginning 67 bp of contig #408 of *wApol* shows conserved similarity with WD0606 (*wmk* gene of *wMel*) (Additional file 14, Figure S8), suggesting wApol may be involved in a male killing *Wolbachia* phenotype.

Discussion

This study highlights diversity of *Wolbachia* harboured by pseudoscorpions. The studied specimens are not closely related, representing different families among pseudoscorpion diversity and contain distinct *Wolbachia*. The small number of pseudoscorpion species studied (3 among 3300 described species) represent only small snapshot of the diversity within this group (3 families among 27 described). This suggests that the diversity of *Wolbachia* among the pseudoscorpion species may likely be underestimated. The species *Geogarypus minor*, a representative of the Geogarypidae family and the species *Chthonius ischnocheles*, a representative of the Chthoniidae family, contain *Wolbachia* closely related and related to supergroup H, present, so far, only in termites. The species *Atemnus politus*, as a representative of the Atemnidae family, contains *Wolbachia* closely related to the *Wolbachia* from the previously described pseudoscorpion *Cordylochernes scorpioides* and appears to form a clade as a sister group to supergroup C (infecting exclusively filarial nematodes) and to supergroup F (infecting filarial nematodes and arthropods). The data suggests that within the evolutionary history of pseudoscorpions, numerous events of horizontal transfers of *Wolbachia* infection have occurred. More detailed, biodiversity analysis of this group will be required to fully understand the evolutionary dynamics of *Wolbachia* within it.

The *Wolbachia* harboured by *Atemnus politus* (*w*Apol) and *Cordylochernes scorpioides* (*w*Csco) form a clade divergent from those *Wolbachia* supergroups described so far (Additional file 15, Figure S9; Additional file 16, Figure S10). The analysis of the *w*Apol draft genome either through phylogenomic

analysis (Figure 5), ANI value or dDDH (Figure 4) also clearly indicates a divergent *Wolbachia* supergroup. It is important to distinguish the notion of “supergroup” from the notion of “species” (which remains problematic for bacteria [40]). The “supergroup” naming only describes the different evolutionary lineages (or clades) of *Wolbachia,* and their limits remain arbitrary (as far as it being a monophyletic group). Thus, we propose that *Wolbachia* from *Atemnus politus* and *Cordylochernes scorpioides* constitute representatives of a new supergroup. *Wolbachia* have currently been assigned into in supergroups A to R. While supergroup R was defined as a clade composed by the endosymbiont from cave spider *Telema* spp. (Wang, Jia et al. 2016), the phylogenetic analysis has been revised and *Wolbachia* from *Telema* spp belong to supergroup A [27]. Our analyses of *ftsZ* and 16S confirm this assignment (Additional file 15, Figure S9; Additional file 16, Figure S10). To avoid any confusion, we propose not to reuse the supergroup R [27] designation and have assigned *Wolbachia* from the pseudoscorpion *Atemnus politus* and *Cordylochernes scorpioides* to novel supergroup S.

As a representative of supergroup S, the draft genome sequence from *wApol*, *Wolbachia* from *Atemnus politus, was* produced. The total length of the *wApol* draft is 1,445,964bp, an average size for a *Wolbachia* from arthropods. The smallest complete *Wolbachia* genome is the symbiont from *Trichogramma pretiosum wTre* which is 1,133,809 bp long [50] and the largest is the symbiont from *Folsomia candida wFol* which is 1,801,626 [11]. The genomes of *Wolbachia* from filarial nematodes are smaller, between 957,990bp and 1,080,084bp [51, 52]. However, the quantitative measures for the assessment of genome assembly using BUSCO indicate that the draft genome *wApol* may be incomplete. Indeed, the percentage of missing single-copy orthologs conserved among proteobacteria and *w*Apol (or missing BUSCOs) (around 22.6%) is close to the percentage observed for smaller genomes as *w*Oo (22.6%; 957,990 bp long genome), *w*Ov (23.1; 960,618 bp long genome) or *w*Ppe (21.7%; 975,127bp long genome), even though this genome is significantly larger. Similarly, the level of IS elements is less than usually observed for most of *Wolbachia* from arthropods (with the exception of *w*Tpre) (Figure 5). Regarding other transposable elements, the *w*Apol genome contains a moderate number of Group II intron-associated genes and mobile elements compared to *Wolbachia* genomes from arthropods, but a high level compared to those of *Wolbachia* from filarial nematodes (Figure 5). Interestingly, it has been suggested that an expansion of mobile genetic elements in bacterial genomes might be a characteristic of initial stages of bacterial adaptation to their host [53]. *w*Apol may represent an intermediate stage between recent acquisition and long-term obligate symbiosis. The presence of these transposable elements may be a contributing factor for the complete *de novo* assembly of *w*Apol. As shown with the *de novo* assembly of *w*AlbB (*Wolbachia* from *Aedes albopictus*), which contains a high number of Group II intron-associated genes [35], a large percentage of long repetitive elements in a genome can limit the efficiency of the *de novo* assembly following hybridization capture protocols.

The annotation of the *w*Apol draft genome was mapped to the KEGG pathway using KASS [54]. Although difficult to conclude about the absence of a particular gene because it is a draft genome, the analysis identifies many conserved pathways (Additional file 17, Table S4; Additional file 18; Table S5) and identifies metabolism pathways present in *w*Apol. For the Vitamin B pathways, it was previously determined that only *w*Cle (*Wolbachia* from bedbug *Cimex lectularius*) contained a complete pathway for

biotin (vitamin B7) and for thiamine (vitamin B1) [55]. More recently, a complete pathway for biotin was identified in *w*Nfla and *w*Nleu (supergroup A) [41] , as well as, *w*Lug and *w*stri (supergroup B) [56]. *w*Apol has the same pathways (with the exception of the gene *bioC*). It was previously demonstrated that *w*Cle provisions biotin, (but not thiamin), significantly contributing to the fitness of host bedbug [55]. In the case of planthoppers symbionts, *w*Lug and *w*stri, *Wolbachia* seems to increase the fecundity of their hosts and it may be related to a beneficial effect of *Wolbachia*-synthesized biotin and riboflavin for the host [56]. The presence of this pathway in *w*Apol suggests it may also be important in the nature of the association between *Wolbachia* and their pseudoscorpion hosts. We identified the complete biotin pathway for *Wolbachia* from the isopod *Armadillidium vulgare w*VulC genome (but not in closely related *w*Con, data not presented), not previously observed. The produced phylogenies of the biotin genes support the concept that the biotin operons may have been acquired by lateral transfer from endosymbiotic bacteria, such as *Cardinium* species, as previously suggested [41, 56]. Our data shows a lack of congruence between the phylogeny of the biotin operons and *Wolbachia* phylogeny, suggesting that this operon might be laterally transferred multiple times and independently along *Wolbachia* evolutionary history.

Another interesting observation concerns the presence of *CydA* and *CydB* genes (cytochrome bd terminal oxidase) which are prokaryotic respiratory quinols. Indeed, these genes are exclusively detected in wApol and studied supergroup A *Wolbachia* (*w*CauA, *w*Mel, wNfla and wNleu). It has been suggested that these genes might be expressed to enhance bacterial tolerance to nitrosative stress and could be a strategy that parasitic bacteria used to escape host immunity systems [57]. The data suggests that these genes might be laterally transfer independently within the supergroup S and supergroup A.

Since the first observation in *Wolbachia* from *Culex pipiens*, the presence of WO bacteriophage in *Wolbachia* has been well documented [58-61]. The observation that they have not been eliminated by a selective pressure suggests that they might provide factors of importance to *Wolbachia* [59]. Not all *Wolbachia* genomes have intact prophage regions; only vestiges of prophage DNA were detected in *Wolbachia* genomes infecting nematodes belonging to C, D or L supergoups [15, 51, 52]. Here, we identified 48 contigs from *wApol* draft which have strong similarities to WO bacteriophage sequences, including partial regions of intact genes, such as structural phage proteins (tail, portal phage, capsid). It is difficult to know if these detected WO regions are fragmented due to selective pressure for elimination or if they are incomplete due to the incompleteness of the draft genome linked to complexity of assembling large repeat sequences after capture-enrichment method. Assembly of the *w*AlbB genome after hybridization capture also failed to assemble the WO region [35]. Despite this, it appears that *w*Apol contains insertion of WO bacteriophage sequences in its genome. Given the number, conservation of protein sequences and diversity of propagic genes (structural and non-structural), an activity of replication and virus particles production seems quite plausible.

Recently, the *cifA* and *cifB* genes contained in the WO region of the Culex *Wolbachia* genome were identified as linked with cytoplasmic incompatibility parasitism [48] and the *wmk* gene was described as candidate gene involved in male-killing [49]. While no orthologues of these genes were identified in *wApol*

for either *cifA* or *cifB*, a small contig was present similarity with *wmk* gene which could be involved in male-killing [49]. Of interest is that for *Wolbachia* from *Cordylochernes scorpioides,* which is very closely related to wApol, antibiotic treatments of *Cordylochernes scorpioides* demonstrate male-killing of the host [33].

Conclusions

Our results emphasize the diversity of *Wolbachia* among the pseudoscorpion family. We identified infection with two different groups of *Wolbachia,* suggesting their independent evolutionary inheritance*,* likely via host-switching. Horizontal transmission of *Wolbachia* among insects has been previously documented [31, 62, 63] but this is the first time that a large group of pseudoscorpions has been analyzed. Among these *Wolbachia*, the symbiont from *Atemnus politus,* was very divergent from previous described supergroups and we propose that a novel clade composed of wApol and wCsco be designated as supergroup S. The produced multi-locus phylogenies and phylogenomics indicate that the supergroup S is closely related to the supergroup F, which contains *Wolbachia* symbionts of arthropods and filarial nematodes and is also closely related to the supergroup C, comprised exclusively *Wolbachia* of filarial nematodes. As the first representative of the supergroup S, we produced a 373-contig draft genome of *w*Apol (1,445,964 bp) using the target enrichment capture method. The *de novo* assembly of wApol may be incomplete, as the presence of large repetitive motifs interferes with complete genome assembly, using this probe hybridization capture method. Nevertheless, supergroup S appears divergent from all the available genomes which were used for probe design [35]. Interestingly, the annotation of the *w*Apol draft genome, even if potentially incomplete, contains the vitamin B biochemical pathway, as opposed to its lack in most of other *Wolbachia*, being more similar to that observed in the *w*Cle symbiont (supergroup F), which has been demonstrated as mutualist with its bedbug host. Of further interest was the finding of a potential orthologue sequence gene of *wmk* gene recently described as as a male-killing gene candidate [49]. The *w*Apol draft genome contains a gene content which might suggest both mutualism and parasitism phenotypes, similar to that of planthopper *Wolbachia* infections (*w*Stri and *w*Lug), characterized by both a cytoplasmic incompatibility phenotype and nutritional mutualism with vitamin B supplementation [56]. The mosaic nature of the pseudoscorpion-*Wolbachia* biology emphasizes the amazing complexity and evolutionary trajectories of these ubiquitous symbionts and provides the background for future comparative studies.

Methods

Source of Material, DNA extraction and Characterization

Specimens were caught in the field using classical soil microfauna recovery methods using a Berlese–Tullgren funnel [64, 65]. The samples were collected at Montpellier in 2017 (43°37'52.0"N 3°52'04.6"E).

DNA samples were extracted using the Monarch® Genomic DNA Purification Kit following the recommended protocol for extraction from tissues (New England Biolabs, USA) with overnight incubation

at 56°C with proteinase K. The quality of the extraction was verified by a PCR targeting the host COI gene (Additional file 17, Table S7). A total of 20 sequences were deposited in the GenBank Data Library: MN923050-MN923069 (Additional file 18, Table S8).

Detection and molecular characterization of *Wolbachia* symbionts

The determination of *Wolbachia* infection in populations was determined by a series of individual specific PCRs (Additional file 17, Table S7). The presence of *Wolbachia* was initially screened using PCR amplification of the *Wolbachia* surface protein (*Wsp*) gene. This pre-screening was further performed by PCR amplification of the cell division protein *FtsZ* gene on 10% of randomly selected specimens. If the results of the two previous markers were not consistent, an amplification of a third marker, the *gatB* gene, was performed.

The molecular characterization of *Wolbachia* was determined by PCR amplification of six genes (16S rDNA gene, *dnaA*, *coxA*, *fbpA*, *gatB* and *ftsZ*) as described in Lefoulon et al. [14] (Additional file 17, Table S7). Nested PCR amplification was necessary to obtain enough PCR products to be directly sequenced. A total of 14 sequences were deposited in the GenBank Data Library: MN931248 to MN931249 and MN931689 to MN931700 (Additional file 19, Table S7).

Next-Generation sequencing of *Atemnus politus*.

Atemnus politus specimens were selected for genomic analysis. Four different libraries preparations were processed: one Illumina and one PacBio library with the *Wolbachia* DNA enrichment protocol, one Illumina and one Pacbio library without enrichment. The enrichment method has been described by Lefoulon et al. [35] and it is based on the use of biotinylated probes to capture *Wolbachia* DNA (probes designed by Roche NimbleGen).

For PacBio sequencing, the Large Enriched Fragment Targeted Sequencing (LEFT-SEQ) method as previously described, was utilized [35] without the DNA fragmentation step. The PacBio library without enrichment was produced using the SMRTbell® Express Template Prep Kit 2.0 for Low DNA input, using the barcoded Overhang Adapter (Kit- 8B, PacBio). The enriched PacBio library was sequenced using 2 SMRT cells with the PacBio RS II system and the library without enrichment used 1 SMRT cell, all on the PacBio Sequel I system.

For Illumina sequencing, an adaptation of the described protocol was performed, eliminating the last steps from PacBio library protocol (end repair, ligation of PacBio adaptor and purification). DNA was fragmented using NEBNext® Ultra™ II FS DNA (NEB, USA) at 20˚C for 30 minutes, resulting in DNA fragments with an average size of 350 bases pairs. We used 100ng of DNA per sample for each capture and used SeqCap barcoded adaptors (Nimblegen, Roche) to process simultaneous multiple samples. The Illumina library without enrichment was produced using the NEBNext® Ultra™ II FS DNA Library Prep Kit following the manufacturer's recommendations (NEB, USA). The enriched Illumina library was sequenced on three independent Ilumina MiSeq runs: one mate-pair 150bp read and two mate-pair 300bp reads. The

unenriched Illumina library was sequenced with one, single-end, 150bp NextSeq run. All sequencing was performed at New England Biolabs.

De novo assembly pipeline

Illumina reads were filtered by quality and "cleaned" using the wrapper Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), and then merged with PEAR (Zhang, Kobert et al. 2014). PacBio circular consensus sequences (CCS) were generated using SMRT® pipe RS_ReadsOfInsert Protocol (PacBio) with a minimum 3 full passes and minimum predicted accuracy superior at 90. The adapters were removed by trimming off the first and last 65 bp of the reads and reads smaller than 20 bp or reads containing residual adaptor sequences (potential chimeric reads) were detected and removed using seqkt (github.com/lh3/seqtk) (analyses were performed with an in-house shell script).

A first hybrid *de novo* assembly was done using Unicycler [66]. The potential contigs belonging to *Wolbachia* were detected by nucleotide similarity using BLASTn [67] and selected. The Illumina reads were mapped against this contig selection using Bowtie2 (with the very sensitive settings) [68] and the PacBio reads using ngmlr (with the PacBio preset settings) [69].

A second hybrid *de novo* assembly was performed with this new selection of reads. Selection of *Wolbachia* contigs was performed using BLASTn a second time. Assembly statistics were calculated using QUAST [70].

In addition, potential contigs belonging to mitochondria of the host were isolated from the first *de novo* assembly using BLASTn against mitochondrial genomes from pseudoscorpion available in the database. Selection of mapped reads were prepared as described above. An assembly of the host mitochondrial genome was performed using Unicycler. The produced mitochondrial sequences was annotated using MITOS (Bernt, Donath et al. 2013).

Comparative genomic analyses and annotation of *wApol*

The comparative genomic analyses described below included analyses of 10 available complete genomes of *Wolbachia and* 4 draft genomes: *w*Mel, *Wolbachia* from *Drosophila Melanogaster* (NC_002978), *w*Cau, *Wolbachia* from *Carposina sasakii* (CP041215) and *w*Nfla, *Wolbachia* from *Nomada flava* (LYUW00000000) for supergroup A; *w*Pip, *Wolbachia* from *Culex quinquefasciatus* (NC_010981), *w*Tpre, *Wolbachia* from *Trichogramma pretiosum* (NZ_CM003641), *w*Lug, *Wolbachia* from *Nilaparvata lugens* and *w*Stri (MUIY01000000), *Wolbachia* from *Laodelphax striatella* (LRUH01000000) for supergroup B; *w*VulC, *Wolbachia* from *Armadillidium vulgare* (ALWU00000000) closely related to the supergroup B; *w*Ppe, *Wolbachia* from *Pratylenchus penetrans* for supergroup L (NZ_MJMG01000000); *w*Cle, *Wolbachia* from *Cimex lectularius* for supergroup F (NZ_AP013028); *w*Fol*, Wolbachia* from *Folsomia candida* for supergroup E (NZ_CP015510); *w*Bm, *Wolbachia* from *Brugia malayi* (NC_006833)

for supergroup D; *w*Ov *Wolbachia* from *Onchocerca volvulus* (NZ_HG810405) and *w*Oo, *Wolbachia* from *Onchocerca ochengi* (NC_018267) from supergroup C.

The Average Nucleotide Identity (ANI) between the *wApol* draft genome and available complete genomes of *Wolbachia* was performed using the ANI Calculator [71]. In addition, an in-silico genome-to-genome comparison was done to calculate a digital DNA-DNA hybridization (DDH) using GGDC [39]. The calculation of dDDH allows analysis of species delineation as alternative to the wet-lab DDH used for current taxonomic techniques. GGDC use a Genome Blast Distance Phylogeny approach (GBDP) to calculated the probability that an intergenomic distance yielded a DDH larger than 70% representing a novel species-delimitation threshold [39]. We used formula 2 to calculate the dDDH because it is more robust using incomplete draft genomes [72]. The completeness of the draft genome was studied using BUSCO v3 [73]. BUSCO estimates the completeness of genomes analyzing gene content and comparing to selection of near-universal single-copy orthologue genes (here, 221 genes in common among proteobacteria (proteobacteria_odb9)).

The processed drafts were analyzed using RAST pipeline [74] and annotated using Prokka [75]. Transposable elements were identified: insertion sequences (ISs) using ISSAGA [76] and group II introns using RAST pipeline. KEGG Orthology (KO) assignment were generated using KASS (KEGG Automatic Annotation Server [54]. KASS assigned orthologue genes by BLAST comparison against KEGG genes database using BBH (bi-directional best hit) method. The same assignment analysis was performed for the *wApol* draft genome, the set of 14 complete or draft genomes and 1 supplementary draft genomes: *w*Nleu, *Wolbachia* from *Nomada eucophthalma* (LYUV00000000). The assigned KOs were ordered in 165 different KEGG pathways (Additional file 20, Table S10). 8 pathways were selected which showed differences between *wApol* annotation with other *Wolbachia* reference and for them, a list of assigned genes involved were identified to study potential losses/acquisition of genes within *Wolbachia* diversity (Additional file 21,Table S11). For some genes of interest, such as the biotin operon, the amino-acid sequences were selected with the KASS assignments and protein orthologues in other *Wolbachia* or bacteria were identified using nblast and then aligned.

The contigs of *wApol* draft genome containing WO bacteriophage region were mapped across the complete genome of phage WOVitA1 (HQ906662). Homologs genes from other *Wolbachia* were then selected using nblast on available data on NCBI. Regarding some genes of interest, homologues of *cifA*, *cifB* and *wmk* were identified in *wApol* using nblast and then mapped with selection of homologues genes from other *Wolbachia*.

Multi-locus phylogeny and Phylogenomics

Two multi-locus phylogenies were performed. The first phylogeny was based on six markers (16S rDNA, *dnaA*, *ftsZ*, *coxA*, *fbpA* and *gatB*), classically used for *Wolbachia* phylogeny [14, 77]. The produced sequences were analyzed with available sequences extracted from 49 *Wolbachia* complete or draft genomes and the addition of sequences from *Wolbachia* from *Zootermopsis angusticollis* and *Zootermopsis nevadensis* (Additional file 19, Table S9). A second phylogeny was based on thirteen

markers (*groEL*, *fabK*, *nuoG*, *NADH dehydrogenase I subunit F*, *aspS*, *gltA*, *coxA*, *ftsZ*, *wsp*, *orpB*, *nuoD*, isocitrate dehydrogenase gene, TPR domain-containing protein gene) which includes less *Wolbachia* strains (because these markers were rarely used for phylogenetic analyses) but included the only *Wolbachia* known to infect pseudoscorpion, *Wolbachia* from the pseudoscorpion *Cordylochernes* scorpioides [33, 38]. This analysis included 14 complete genomes in addition to the *w*Apol draft genome and *w*Csco *Wolbachia* from *Cordylochernes scorpioides* (Additional file 19, Table S9)

For phylogenomic analyses, single-copy orthologue genes between a selection of *Wolbachia* genomes were identified using Orthofinder [78]. Two types of phylogenomics studies were performed: one included only 15 complete genomes and one included 31 complete or draft genomes (Additional file 19, Table S9). Variations in completeness of draft genomes can be variable and have a negative effect on the robustness of the analyses, and thus the two separate analyses were performed.

The orthologue sequence alignments were generated by MAFFT [79]. For the multi-locus phylogenies, a supermatrix of these six alignments was generated using Seaview [80], and for phylogenomics, the supermatrix was produced by Orthofinder (implemented as functionality). For the later, the poorly aligned positions of the produced orthologue genes alignments were eliminated using Gblocks [81]. The phylogenetic analyses were performed with Maximum Likelihood inference using IQ-TREE [82]. The most appropriate model of evolution was evaluated by Modelfinder [83] (implemented as functionality of IQ-TREE). The robustness of each node was evaluated by a bootstrap test (1,000 replicates). The phylogenetic trees were edited by FigTree (https://github.com/rambaut/figtree/) and Inkscape (https://inkscape.org/).

Declarations

Ethics Approval and consent to participate

Not Applicable

Consent for Publication

Not Applicable

Availability of Data and Materials

Data generated are available in GenBank: BioProject SUB6649337; BioSample SAMN13481355 for *Wolbachia* endosymbiont strain of *Atemnus* sp. (genome: WQMQ00000000). The raw data are available in GenBank as Sequence Read Archive (SRA): SRR10881598 and SRR10881597. In addition, a total of 34

sequences were deposited in the GenBank: MN923050 to MN923069, MN931248 to MN931249 and MN931689 to MN931700.

Competing Interests

EL and BS are employed by New England Biolabs, Inc., who provided funding for this project.

Funding

Funded by Internal Research Funds from New England Bioalbs, Inc. to BES.

Author contributions

EL, LG and BES conceived and designed the experiments. EL, TC, MPS and FB performed the experiments. EL and LG analyzed the data. LG and CM contributed materials. EL, LG and BES wrote the main manuscript text. All authors reviewed the manuscript.

Acknowledgments

We thank Mark Judson from MNHN for the confirmation of the morphological identification of the pseudoscorpion specimens. We thank Andy Gardner, Tom Evans, Rich Roberts, Don Comb and Jim Ellard from New England Biolabs for their support. We thank Laurie Mazzola, Danielle Fuchs and and Kristen Augulewicz from the NEB Sequencing Core.

Author Affiliations

Emilie Lefoulon:

Molecular Parasitology Group, Molecular Enzymes Division New England Biolabs, Inc.,

Ipswich, United States of America;

email: lefoulon.emilie@gmailcom

Travis Clark

Molecular Parasitology Group, Molecular Enzymes Division New England Biolabs, Inc.,

Ipswich, United States of America

email: travismichaelclark@gmail.com

Fanni Borveto

MIVEGEC - UMR 5290

39 Avenue Charles Flahault,

34090 Montpellier, France

email: fanni.borveto@gmail.com

Marco Perriat-Sanguinet

Université de Montpellier – C/C 064
Place Eugène Bataillon
34095 Montpellier cedex 05

ISEM, University of Montpellier, Montpellier, France

email: marco.perriat-sanguinet@umontpellier.fr

Catherine Moulia

Université de Montpellier – C/C 064
Place Eugène Bataillon
34095 Montpellier cedex 05

ISEM, University of Montpellier, Montpellier, France

email: catherine.moulia@umontpellier.fr

Barton E. Slatko

Molecular Parasitology Group, Molecular Enzymes Division New England Biolabs, Inc.,

Ipswich, United States of America

email: slatko@neb.com

Laurent Gavotte

ISEM - UMR 5554
Université de Montpellier – C/C 064
Place Eugène Bataillon
34095 Montpellier cedex 05

ISEM, University of Montpellier, Montpellier, France

email: laurent.gavotte@umontpellier.fr

References

1. Sironi M, Bandi C, Sacchi L, Di Sacco B, Damiani G, Genchi C: **Molecular evidence for a close relative of the arthropod endosymbiont Wolbachia in a filarial worm**. *Molecular and biochemical parasitology* 1995, **74**(2):223-227.
2. Werren JH, Zhang W, Guo LR: **Evolution and phylogeny of Wolbachia: reproductive parasites of arthropods**. *Proceedings Biological Sciences* 1995, **261**(1360):55-63.
3. Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH: **How many species are infected with Wolbachia?–A statistical analysis of current data**. *FEMS microbiology letters* 2008, **281**(2):215-220.
4. Zug R, Hammerstein P: **Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected**. *PloS one* 2012, **7**(6):e38544.
5. Sazama EJ, Ouellette SP, Wesner JS: **Bacterial Endosymbionts Are Common Among, but not Necessarily Within, Insect Species**. *Environ Entomol* 2019, **48**(1):127-133.
6. Martin C, Gavotte L: **The bacteria Wolbachia in filariae, a biological Russian dolls' system: new trends in antifilarial treatments**. *Parasite* 2010, **17**(2):79-89.
7. Hoerauf A, Mand S, Volkmann L, Buttner M, Marfo-Debrekyei Y, Taylor M, Adjei O, Buttner DW: **Doxycycline in the treatment of human onchocerciasis: Kinetics of Wolbachia endobacteria reduction and of inhibition of embryogenesis in female Onchocerca worms**. *Microbes and infection / Institut Pasteur* 2003, **5**(4):261-273.
8. Werren JH, Baldo L, Clark ME: **Wolbachia: master manipulators of invertebrate biology**. *Nature reviews Microbiology* 2008, **6**(10):741-751.
9. Zug R, Hammerstein P: **Bad guys turned nice? A critical assessment of Wolbachia mutualisms in arthropod hosts**. *Biol Rev Camb Philos Soc* 2015, **90**(1):89-111.
10. Comandatore F, Cordaux R, Bandi C, Blaxter M, Darby A, Makepeace BL, Montagna M, Sassera D: **Supergroup C Wolbachia, mutualist symbionts of filarial nematodes, have a distinct genome structure**. *Open Biol* 2015, **5**(12):150099.

11. Kampfraath AA, Klasson L, Anvar SY, Vossen R, Roelofs D, Kraaijeveld K, Ellers J: **Genome expansion of an obligate parthenogenesis-associated Wolbachia poses an exception to the symbiont reduction model**. *BMC Genomics* 2019, **20**(1):106.
12. Bandi C, Anderson TJ, Genchi C, Blaxter ML: **Phylogeny of *Wolbachia* in filarial nematodes**. *Proceedings Biological Sciences* 1998, **265**(1413):2407-2413.
13. Casiraghi M, Bain O, Guerrero R, Martin C, Pocacqua V, Gardner SL, Franceschi A, Bandi C: **Mapping the presence of Wolbachia pipientis on the phylogeny of filarial nematodes: evidence for symbiont loss during evolution**. *International journal for parasitology* 2004, **34**(2):191-203.
14. Lefoulon E, Bain O, Makepeace BL, d'Haese C, Uni S, Martin C, Gavotte L: **Breakdown of coevolution between symbiotic bacteria Wolbachia and their filarial hosts**. *PeerJ* 2016, **4**:e1840.
15. Brown AM, Wasala SK, Howe DK, Peetz AB, Zasada IA, Denver DR: **Genomic evidence for plant-parasitic nematodes as the earliest Wolbachia hosts**. *Sci Rep* 2016, **6**:34955.
16. Haegeman A, Vanholme B, Jacob J, Vandekerckhove TT, Claeys M, Borgonie G, Gheysen G: **An endosymbiotic bacterium in a plant-parasitic nematode: member of a new Wolbachia supergroup**. *International journal for parasitology* 2009, **39**(9):1045-1054.
17. Ferri E, Bain O, Barbuto M, Martin C, Lo N, Uni S, Landmann F, Baccei SG, Guerrero R, de Souza Lima S *et al*: **New insights into the evolution of Wolbachia infections in filarial nematodes inferred from a large range of screened species**. *PloS one* 2011, **6**(6):e20843.
18. Lefoulon E, Gavotte L, Junker K, Barbuto M, Uni S, Landmann F, Laaksonen S, Saari S, Nikander S, de Souza Lima S *et al*: **A new type F *Wolbachia* from Splendidofilariinae (Onchocercidae) supports the recent emergence of this supergroup**. *International journal for parasitology* 2012, **42**(11):1025-1036.
19. Bordenstein S, Rosengaus RB: **Discovery of a novel Wolbachia super group in Isoptera**. *Current microbiology* 2005, **51**(6):393-398.
20. Glowska E, Dragun-Damian A, Dabert M, Gerth M: **New Wolbachia supergroups detected in quill mites (Acari: Syringophilidae)**. *Infect Genet Evol* 2015, **30**:140-146.
21. Lo N, Casiraghi M, Salati E, Bazzocchi C, Bandi C: **How many wolbachia supergroups exist?** *Molecular biology and evolution* 2002, **19**(3):341-346.
22. Lo N, Paraskevopoulos C, Bourtzis K, O'Neill SL, Werren JH, Bordenstein SR, Bandi C: **Taxonomic status of the intracellular bacterium Wolbachia pipientis**. *International journal of systematic and evolutionary microbiology* 2007, **57**(Pt 3):654-657.
23. Ros VI, Fleming VM, Feil EJ, Breeuwer JA: **How diverse is the genus Wolbachia? Multiple-gene sequencing reveals a putatively new Wolbachia supergroup recovered from spider mites (Acari: Tetranychidae)**. *Applied and environmental microbiology* 2009, **75**(4):1036-1043.
24. Rowley SM, Raven RJ, McGraw EA: **Wolbachia pipientis in Australian spiders**. *Current microbiology* 2004, **49**(3):208-214.
25. Wang GH, Jia LY, Xiao JH, Huang DW: **Discovery of a new Wolbachia supergroup in cave spider species and the lateral transfer of phage WO among distant hosts**. *Infect Genet Evol* 2016, **41**:1-7.

26. Baldo L, Werren JH: **Revisiting Wolbachia supergroup typing based on WSP: spurious lineages and discordance with MLST**. *Current microbiology* 2007, **55**(1):81-87.
27. Gerth M: **Classification of Wolbachia (Alphaproteobacteria, Rickettsiales): No evidence for a distinct supergroup in cave spiders**. *Infect Genet Evol* 2016, **43**:378-380.
28. Almerao MP, Fagundes NJ, de Araujo PB, Verne S, Grandjean F, Bouchon D, Araujo AM: **First record of Wolbachia in South American terrestrial isopods: Prevalence and diversity in two species of Balloniscus (Crustacea, Oniscidea)**. *Genet Mol Biol* 2012, **35**(4 (suppl)):980-989.
29. Bouchon D, Rigaud T, Juchault P: **Evidence for widespread Wolbachia infection in isopod crustaceans: molecular identification and host feminization**. *Proceedings Biological Sciences* 1998, **265**(1401):1081-1090.
30. Badawi M, Moumen B, Giraud I, Greve P, Cordaux R: **Investigating the Molecular Genetic Basis of Cytoplasmic Sex Determination Caused by Wolbachia Endosymbionts in Terrestrial Isopods**. *Genes (Basel)* 2018, **9**(6).
31. Baldo L, Ayoub NA, Hayashi CY, Russell JA, Stahlhut JK, Werren JH: **Insight into the routes of Wolbachia invasion: high levels of horizontal transfer in the spider genus Agelenopsis revealed by Wolbachia strain and mitochondrial DNA diversity**. *Molecular ecology* 2008, **17**(2):557-569.
32. Baldo L, Prendini L, Corthals A, Werren JH: **Wolbachia are present in southern african scorpions and cluster with supergroup F**. *Current microbiology* 2007, **55**(5):367-373.
33. Zeh DW, Zeh JA, Bonilla MM: **Wolbachia, sex ratio bias and apparent male killing in the harlequin beetle riding pseudoscorpion**. *Heredity (Edinb)* 2005, **95**(1):41-49.
34. Geniez S, Foster JM, Kumar S, Moumen B, Leproust E, Hardy O, Guadalupe M, Thomas SJ, Boone B, Hendrickson C *et al*: **Targeted genome enrichment for efficient purification of endosymbiont DNA from host DNA**. *Symbiosis* 2012, **58**(1-3):201-207.
35. Lefoulon E, Vaisman N, Frydman HM, Sun L, Foster JM, Slatko BE: **Large Enriched Fragment Targeted Sequencing (LEFT-SEQ) Applied to Capture of Wolbachia Genomes**. *Sci Rep* 2019, **9**(1):5939.
36. Gardini G, Galli L, Zinni M: **Redescription of Geogarypus minor, type species of the genus Geogarypus, and description of a new species from Italy (Pseudoscorpiones: Geogarypidae)**. *Journal of Arachnology* 2017, **45**(3):424-443.
37. Kotrbova J, Opatova V, Gardini G, Stahlavsky F: **Karyotype diversity of pseudoscorpions of the genus Chthonius (Pseudoscorpiones, Chthoniidae) in the Alps**. *Comp Cytogenet* 2016, **10**(3):325-345.
38. Bordenstein SR, Paraskevopoulos C, Dunning Hotopp JC, Sapountzis P, Lo N, Bandi C, Tettelin H, Werren JH, Bourtzis K: **Parasitism and mutualism in Wolbachia: what the phylogenomic trees can and cannot say**. *Molecular biology and evolution* 2009, **26**(1):231-241.
39. Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M: **Genome sequence-based species delimitation with confidence intervals and improved distance functions**. *BMC Bioinformatics* 2013, **14**:60.
40. Chung M, Munro JB, Tettelin H, Dunning Hotopp JC: **Using Core Genome Alignments To Assign Bacterial Species**. *mSystems* 2018, **3**(6).

41. Gerth M, Bleidorn C: **Comparative genomics provides a timeframe for Wolbachia evolution and exposes a recent biotin synthesis operon transfer**. *Nat Microbiol* 2016, **2**:16241.
42. Pichon S, Bouchon D, Cordaux R, Chen L, Garrett RA, Greve P: **Conservation of the Type IV secretion system throughout Wolbachia evolution**. *Biochem Biophys Res Commun* 2009, **385**(4):557-562.
43. Rances E, Voronin D, Tran-Van V, Mavingui P: **Genetic and functional characterization of the type IV secretion system in Wolbachia**. *Journal of bacteriology* 2008, **190**(14):5020-5030.
44. Singhal K, Mohanty S: **Genome organisation and comparative genomics of four novel Wolbachia genome assemblies from Indian Drosophila host**. *Funct Integr Genomics* 2019, **19**(4):617-632.
45. Ramirez-Puebla ST, Ormeno-Orrillo E, Vera-Ponce de Leon A, Lozano L, Sanchez-Flores A, Rosenblueth M, Martinez-Romero E: **Genomes of Candidatus Wolbachia bourtzisii wDacA and Candidatus Wolbachia pipientis wDacB from the Cochineal Insect Dactylopius coccus (Hemiptera: Dactylopiidae)**. *G3 (Bethesda)* 2016, **6**(10):3343-3349.
46. Sinha A, Li Z, Sun L, Carlow CKS: **Complete Genome Sequence of the Wolbachia wAlbB Endosymbiont of Aedes albopictus**. *Genome Biol Evol* 2019, **11**(3):706-720.
47. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS: **PHASTER: a better, faster version of the PHAST phage search tool**. *Nucleic Acids Res* 2016, **44**(W1):W16-21.
48. LePage DP, Metcalf JA, Bordenstein SR, On J, Perlmutter JI, Shropshire JD, Layton EM, Funkhouser-Jones LJ, Beckmann JF, Bordenstein SR: **Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility**. *Nature* 2017, **543**(7644):243-247.
49. Perlmutter JI, Bordenstein SR, Unckless RL, LePage DP, Metcalf JA, Hill T, Martinez J, Jiggins FM, Bordenstein SR: **The phage gene wmk is a candidate for male killing by a bacterial endosymbiont**. *PLoS Pathog* 2019, **15**(9):e1007936.
50. Lindsey AR, Werren JH, Richards S, Stouthamer R: **Comparative Genomics of a Parthenogenesis-Inducing Wolbachia Symbiont**. *G3 (Bethesda)* 2016, **6**(7):2113-2123.
51. Darby AC, Armstrong SD, Bah GS, Kaur G, Hughes MA, Kay SM, Koldkjaer P, Radford AD, Blaxter ML, Tanya VN *et al*: **Analysis of gene expression from the Wolbachia genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis**. *Genome research* 2012:doi: 10.1242/bio.2012737.
52. Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, Bhattacharyya A, Kapatral V, Kumar S, Posfai J *et al*: **The Wolbachia genome of Brugia malayi: endosymbiont evolution within a human pathogenic nematode**. *PLoS biology* 2005, **3**(4):e121.
53. McCutcheon JP, Moran NA: **Extreme genome reduction in symbiotic bacteria**. *Nature reviews Microbiology* 2011, **10**(1):13-26.
54. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M: **KAAS: an automatic genome annotation and pathway reconstruction server**. *Nucleic Acids Res* 2007, **35**(Web Server issue):W182-185.
55. Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T: **Evolutionary origin of insect-Wolbachia nutritional mutualism**. *Proc Natl Acad Sci U S A* 2014, **111**(28):10257-10262.

56. Ju JF, Bing XL, Zhao DS, Guo Y, Xi Z, Hoffmann AA, Zhang KJ, Huang HJ, Gong JT, Zhang X *et al*: **Wolbachia supplement biotin and riboflavin to enhance reproduction in planthoppers**. *ISME J* 2019.
57. Giuffre A, Borisov VB, Arese M, Sarti P, Forte E: **Cytochrome bd oxidase and bacterial tolerance to oxidative and nitrosative stress**. *Biochim Biophys Acta* 2014, **1837**(7):1178-1187.
58. Fujii Y, Kubo T, Ishikawa H, Sasaki T: **Isolation and characterization of the bacteriophage WO from Wolbachia, an arthropod endosymbiont**. *Biochem Biophys Res Commun* 2004, **317**(4):1183-1188.
59. Gavotte L, Henri H, Stouthamer R, Charif D, Charlat S, Bouletreau M, Vavre F: **A Survey of the bacteriophage WO in the endosymbiotic bacteria Wolbachia**. *Molecular biology and evolution* 2007, **24**(2):427-435.
60. Masui S, Kuroiwa H, Sasaki T, Inui M, Kuroiwa T, Ishikawa H: **Bacteriophage WO and virus-like particles in Wolbachia, an endosymbiont of arthropods**. *Biochem Biophys Res Commun* 2001, **283**(5):1099-1104.
61. Wright JD, Sjostrand FS, Portaro JK, Barr AR: **The ultrastructure of the rickettsia-like microorganism Wolbachia pipientis and associated virus-like bodies in the mosquito Culex pipiens**. *J Ultrastruct Res* 1978, **63**(1):79-85.
62. O'Neill SL, Giordano R, Colbert AM, Karr TL, Robertson HM: **16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects**. *Proc Natl Acad Sci U S A* 1992, **89**(7):2699-2702.
63. Gerth M, Rothe J, Bleidorn C: **Tracing horizontal Wolbachia movements among bees (Anthophila): a combined approach using multilocus sequence typing data and host phylogeny**. *Molecular ecology* 2013, **22**(24):6149-6162.
64. Kaydan MB, Benedicty ZK, Kiss B, Szita E: **A survey of scale insects in soil samples from Europe (Hemiptera, Coccomorpha)**. *Zookeys* 2016(565):1-28.
65. Sabu TK, Shiju RT, Vinod K, Nithya S: **A comparison of the pitfall trap, Winkler extractor and Berlese funnel for sampling ground-dwelling arthropods in tropical montane cloud forests**. *J Insect Sci* 2011, **11**:28.
66. Wick RR, Judd LM, Gorrie CL, Holt KE: **Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads**. *PLoS Comput Biol* 2017, **13**(6):e1005595.
67. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST+: architecture and applications**. *BMC Bioinformatics* 2009, **10**:421.
68. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2**. *Nat Methods* 2012, **9**(4):357-359.
69. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC: **Accurate detection of complex structural variations using single-molecule sequencing**. *Nat Methods* 2018, **15**(6):461-468.
70. Gurevich A, Saveliev V, Vyahhi N, Tesler G: **QUAST: quality assessment tool for genome assemblies**. *Bioinformatics* 2013, **29**(8):1072-1075.

71. Yoon SH, Ha SM, Lim J, Kwon S, Chun J: **A large-scale evaluation of algorithms to calculate average nucleotide identity**. *Antonie Van Leeuwenhoek* 2017, **110**(10):1281-1286.
72. Auch AF, Klenk HP, Goker M: **Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs**. *Stand Genomic Sci* 2010, **2**(1):142-148.
73. Seppey M, Manni M, Zdobnov EM: **BUSCO: Assessing Genome Assembly and Annotation Completeness**. *Methods Mol Biol* 2019, **1962**:227-245.
74. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M *et al*: **The RAST Server: rapid annotations using subsystems technology**. *BMC Genomics* 2008, **9**:75.
75. Seemann T: **Prokka: rapid prokaryotic genome annotation**. *Bioinformatics* 2014, **30**(14):2068-2069.
76. Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M: **ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes**. *Genome Biol* 2011, **12**(3):R30.
77. Baldo L, Dunning Hotopp JC, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, Hayashi C, Maiden MC, Tettelin H, Werren JH: **Multilocus sequence typing system for the endosymbiont Wolbachia pipientis**. *Applied and environmental microbiology* 2006, **72**(11):7098-7110.
78. Emms DM, Kelly S: **OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy**. *Genome Biol* 2015, **16**:157.
79. Katoh K, Misawa K, Kuma K, Miyata T: **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform**. *Nucleic Acids Res* 2002, **30**(14):3059-3066.
80. Gouy M, Guindon S, Gascuel O: **SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building**. *Molecular biology and evolution* 2010, **27**(2):221-224.
81. Talavera G, Castresana J: **Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments**. *Syst Biol* 2007, **56**(4):564-577.
82. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ: **IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies**. *Molecular biology and evolution* 2015, **32**(1):268-274.
83. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS: **ModelFinder: fast model selection for accurate phylogenetic estimates**. *Nat Methods* 2017, **14**(6):587-589.

Additional Files (Supplementary Information)

Additional file 1: Figure S1 Representation of mitochondrial genomes of three pseudoscorpions: *Atemnus politus*, *Paratemnoides elongatus* and *Pseudogarypus banksi*.

Additional file 2: Figure S2, Unrooted phylogenetic trees of *Wolbachia* based on 13 markers by Maximum Likelihood.

Addtional file 3: Figure S3, BUSCO Assessment Result of 15 draft or complete *Wolbachia* genomes.

Additional file 4: Table S1, Summary of BUSCOs analyses of *Wolbachia* complete genomes and *w*Apol.

Additional file 5: Table S2, List of potential insertion sequences elements (ISs).

Additional file 6: Table S3, List of Group II intron-associated genes.

Additional file 7: Table S4, List of mobile element genes.

Additional file 8: Figure S4, Diagram of the contigs of *w*Apol draft genome containing the effector type IV secretion system (T4SSs) genes.

Additional file 9: Table S5, List of coding sequences (cds) involved in the Vir-like type IV secretion system (T4SS).

Additional file 10: Table S6, Contigs of *w*Apol identified as homolog with WO bacteriophage.

Additional file 11: Figure S5, Phylogeny of homologues of WOVitA1 bacteriophage region (gwv 1154, gwv 1152, gwv 1140 to 1142, gwv 1123 to 1127).

Additional file 12: Figure S6, Phylogeny of homologues of WOVitA1 bacteriophage region (gwv 1119 to 1121, gwv 1115 to 1116, gwv 1114).

Additional file 13: Figure S7, Phylogeny of homologues of WOVitA1 bacteriophage region (gwv 1103 to 1113, gwv 1096, gwv 1093 to 1094).

Additional file 14: Figure S8, Phylogeny of homologues of the *wmk* gene.

Additional file 15: Figure S9, Phylogeny of *Wolbachia* based on 16S gene.

Additional file 16: Figure S10, Phylogeny of *Wolbachia* based on ftsZ gene.

Additional file 17: Table S7, PCR primers and conditions used in this study.

Additional file 18: Table S8, NCBI accession of CO1 sequences of pseudoscorpion produced in the study.

Additional file 19: Table S9, NCBI accession of sequences of *Wolbachia* used in the study and detail of sequences included in molecular analyses.

Additional file 20: Table S10, Summary of number of genes assigned to 165 different KEGG pathways from *Wolbachia* genomes using KASS.

Additional file 21: Table S11, Details of genes assigned to selection of KEGG pathways from *Wolbachia* genomes using KASS.

Figures

Figure 1

Phylogenetic tree of pseudoscorpions based on COI. The total length of datasets is 659 bp. The topology was inferred using Maximum Likelihood (ML) inference using IQTREE. The Best-fit model calculated using ModelFinder according to BIC index was TVM+R7. Nodes are associated with Bootstrap values based on 1,000 replicates, only bootstrap value superior to 70 are indicated.

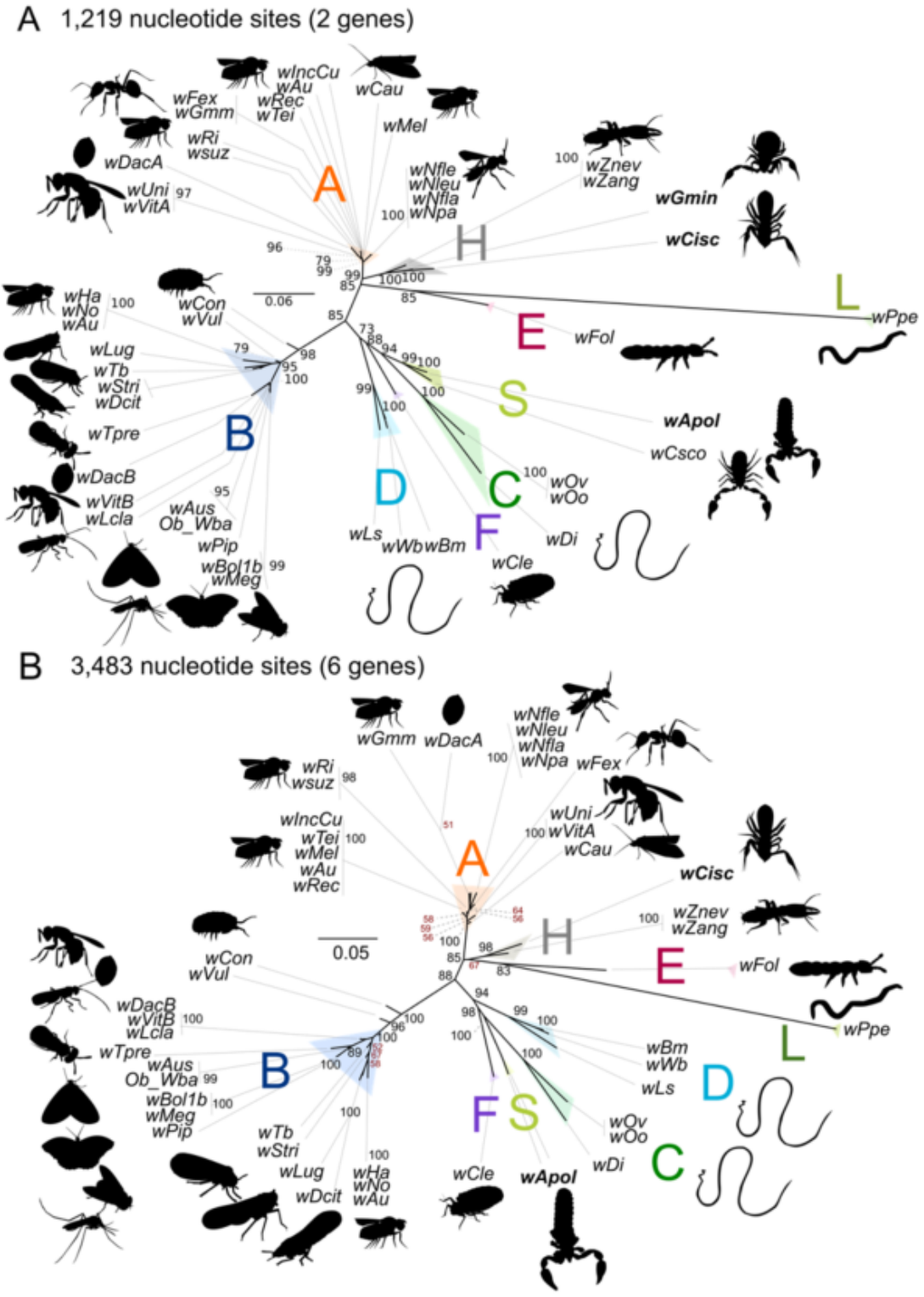

Figure 2

Unrooted phylogenetic trees of Wolbachia based on 2 and 6 markers by Maximum Likelihood. A) Analysis based on concatenation of ftsZ and coxA; the total length of datasets is 1,219 bp. The topology was inferred using Maximum Likelihood (ML) inference using IQTREE. The Best-fit model calculated using ModelFinder according to BIC index was TIM3+I+G4. B) Analysis based on concatenation of 16S rDNA, dnaA, ftsZ, coxA, fbpA and gatB; the total length of datasets is 3,483 bp. The Best-fit model calculated using ModelFinder according to BIC index was TPM3u+I+G4. Nodes are associated with Bootstrap values based on 1,000 replicates, only bootstrap value superior to 70 are indicated. The Wolbachia supergroups (A–S) are indicated.

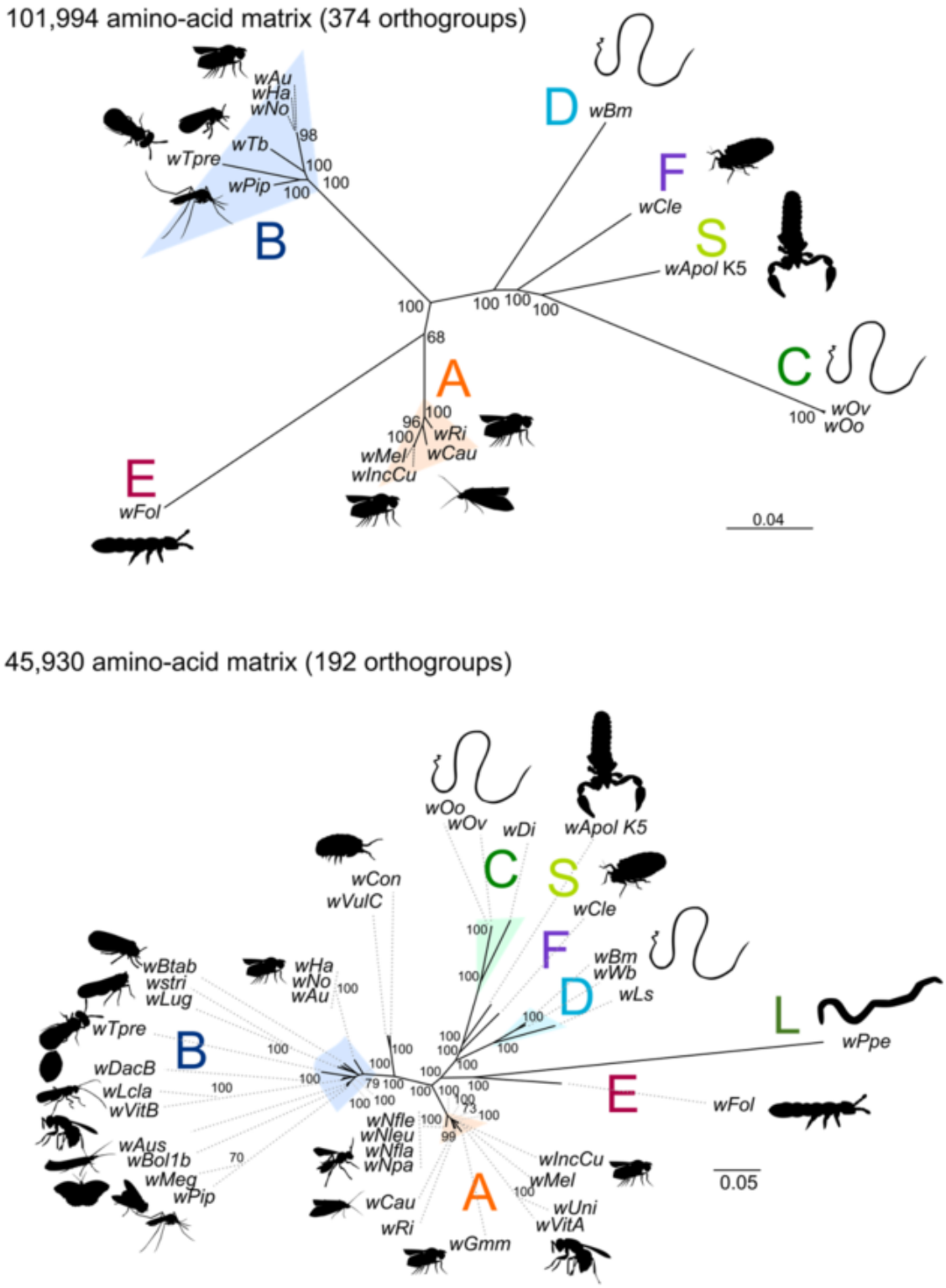

Figure 3

Phylogenomics analyses of Wolbachia. A) Analysis based on concatenation of 374 single-copy orthogroups representing 101,994 amino-acid matrix. The topology was inferred using Maximum Likelihood (ML) inference using IQTREE. The Best-fit model calculated using ModelFinder according to BIC index was HIVb+F+R4 B) Analysis based on concatenation of 192 single-copy orthogroups representing 45,930 amino-acid matrix. The Best-fit model calculated using ModelFinder according to BIC

index was HIVb+F+R4. Nodes are associated with Bootstrap values based on 1,000 replicates, only bootstrap value superior to 70 are indicated. The Wolbachia supergroups (A–S) are indicated.

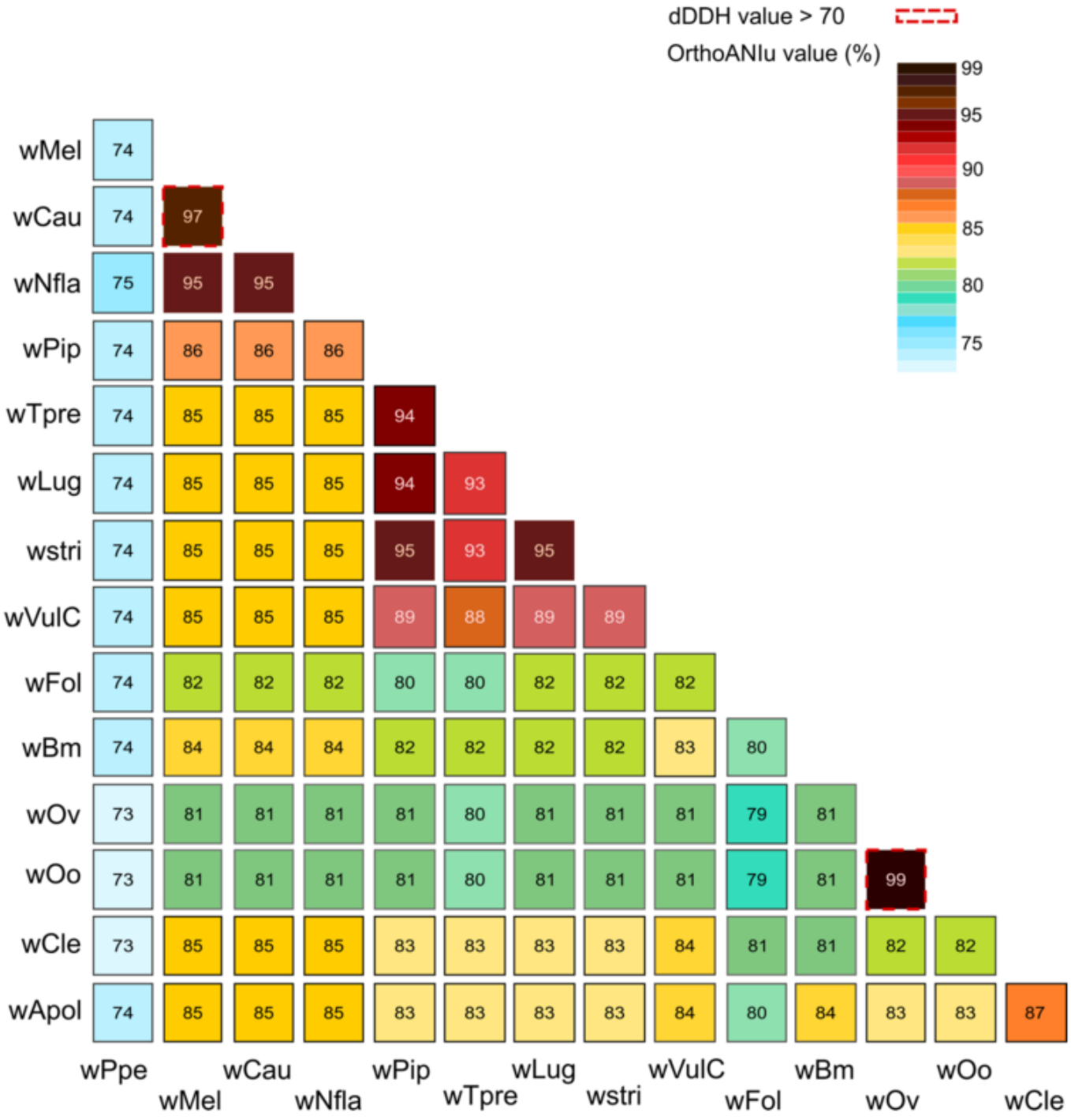

Figure 4

Summary of ANI and dDDH calculation of Wolbachia genome.

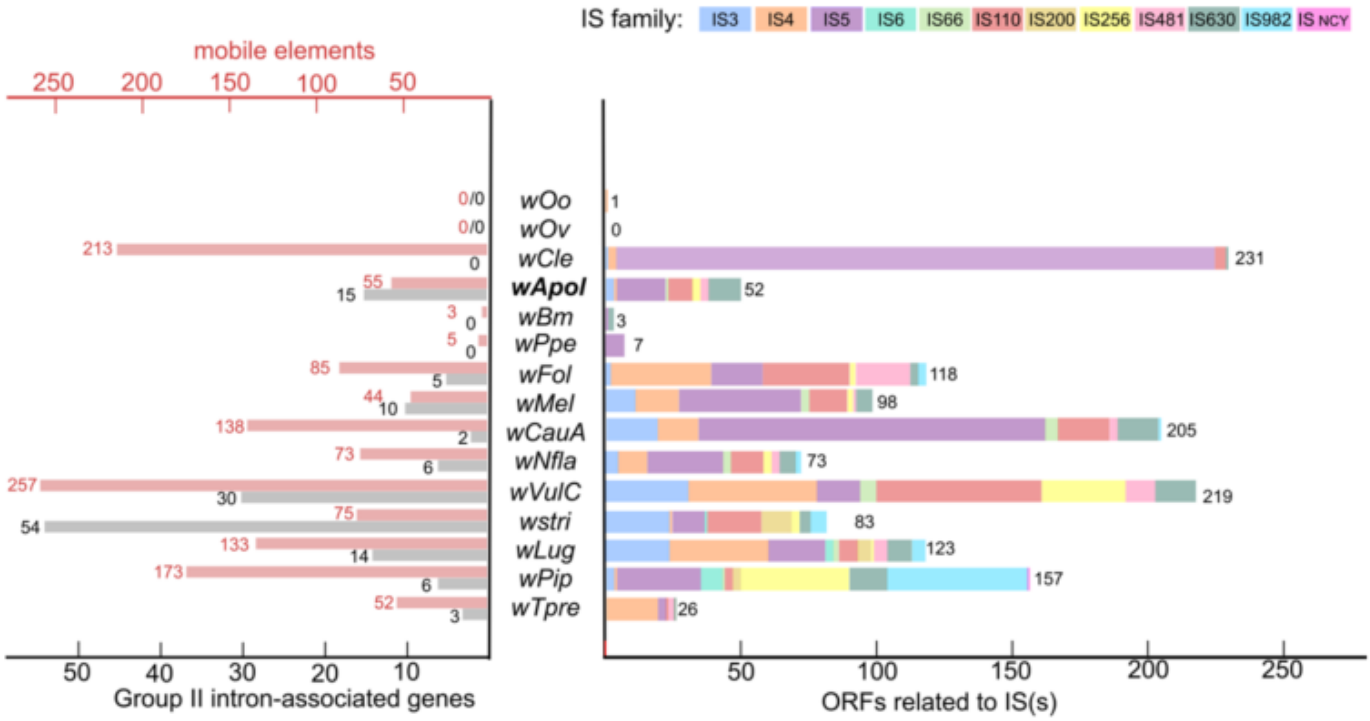

Figure 5

Graph of IS elements and group II intron-associated genes identified in Wolbachia genomes. Transposable elements were identified: insertion sequences (ISs) using ISSAGA and group II introns and mobile elements protein using RAST pipeline.

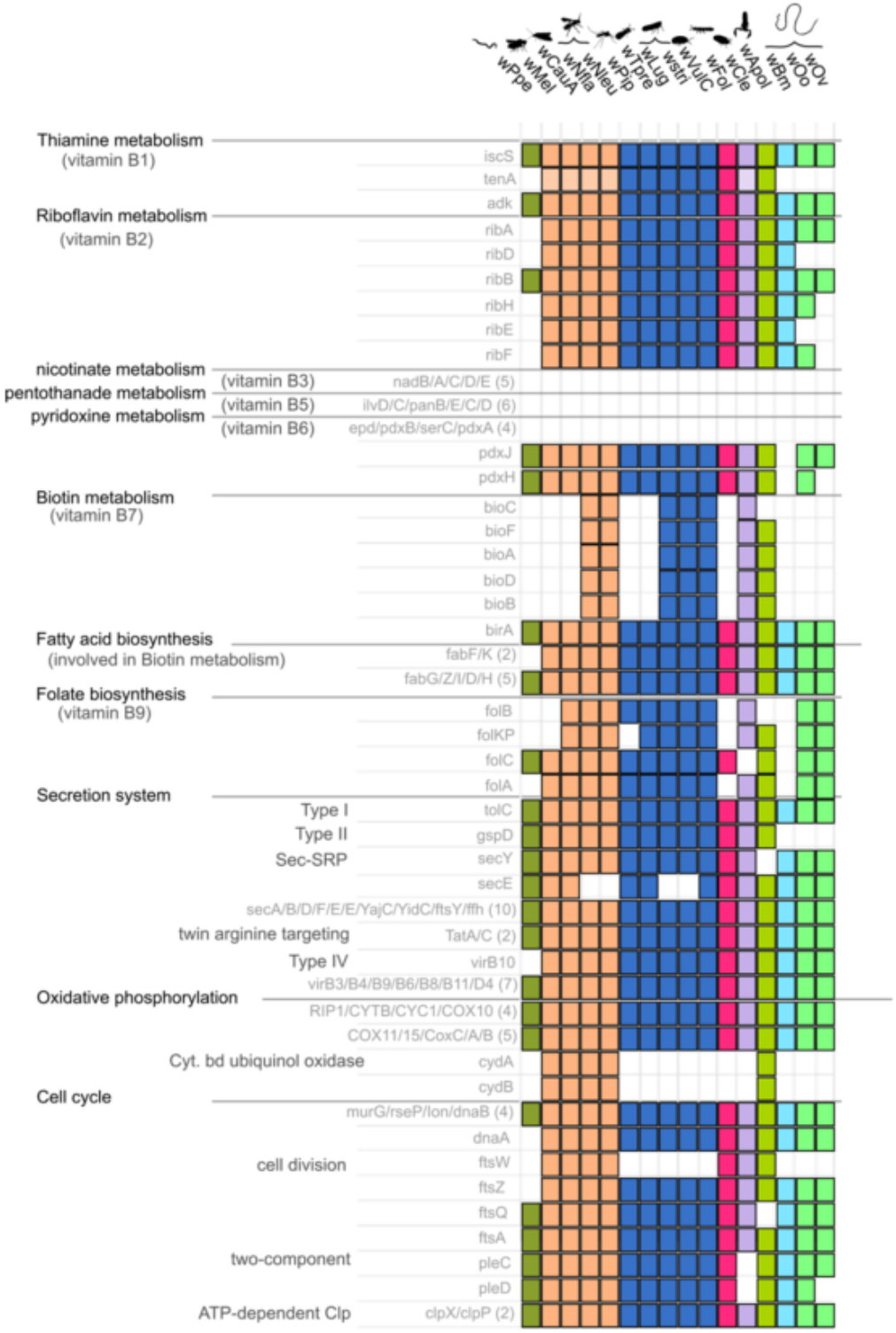

Figure 6

Summary of the metabolic pathways detected in Wolbachia genomes, using KASS.

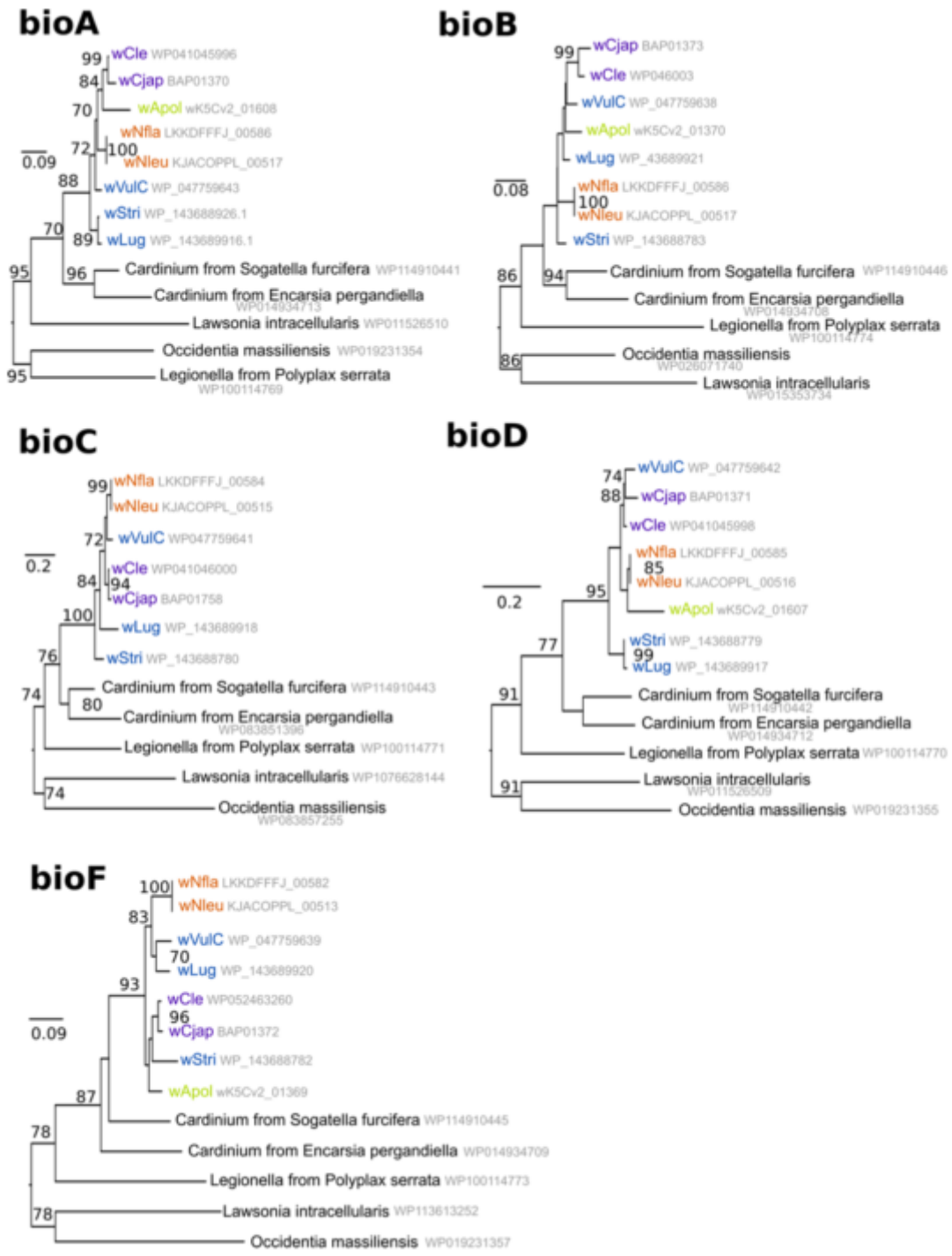

Figure 7

Phylogenetic analysis of biotin (bioA, bioB, bioC, bioD and bioF) synthesis genes of Wolbachia. The length of datasets is 441 amino-acid for bioA, 315 aa for bioB, 255 aa for bioC, 218 aa for bioC and 381 aa for bioF. The topology was inferred using Maximum Likelihood (ML) inference using IQTREE. The Best-fit model calculated using ModelFinder according to BIC index were: cpREV+I+G4 for bioA, FLU+G4 for bioB, JTT+F+G4 for bioC, cpREV+G4 for bioD and cpREV+G4 for bioF. Nodes are associated with Bootstrap values based on 1,000 replicates, only bootstrap value superior to 70 are indicated. The

Wolbachia supergroups are indicated by the color: purple for supergroup F; dark blue for supergroup A; orange for supergroup A and green for supergroup S.

Supplementary Files

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

- Additionalfile7TableS4.xlsx
- Additionalfile4TableS1.xlsx
- Additionalfile10TableS6.xlsx
- Additionalfile18TableS8.xlsx
- Additionalfile20TableS10.xlsx
- Additionalfile9TableS5.xlsx
- Additionalfile21TableS11.xlsx
- Additionalfile17TableS7.xlsx
- Additionalfile19TableS9.xlsx
- Additionalfile6TableS3.xlsx
- Additionalfile5TableS2.xlsx
- Supplementaryfigures.docx