

Diminutive, degraded but dissimilar: *Wolbachia* genomes from filarial nematodes do not conform to a single paradigm

Emilie Lefoulon^{1†}, Travis Clark¹, Ricardo Guerrero², Israel Cañizales^{2,3}, Jorge Manuel Cardenas-Callirgos⁴, Kerstin Junker⁵, Nathaly Vallarino-Lhermitte⁶, Benjamin L. Makepeace⁷, Alistair C. Darby⁷, Jeremy M. Foster¹, Coralie Martin⁶ and Barton E. Slatko^{1,*}

Abstract

Wolbachia are alpha-proteobacteria symbionts infecting a large range of arthropod species and two different families of nematodes. Interestingly, these endosymbionts are able to induce diverse phenotypes in their hosts: they are reproductive parasites within many arthropods, nutritional mutualists within some insects and obligate mutualists within their filarial nematode hosts. Defining *Wolbachia* 'species' is controversial and so they are commonly classified into 17 different phylogenetic lineages, termed supergroups, named A–F, H–Q and S. However, available genomic data remain limited and not representative of the full *Wolbachia* diversity; indeed, of the 24 complete genomes and 55 draft genomes of *Wolbachia* available to date, 84% belong to supergroups A and B, exclusively composed of *Wolbachia* from arthropods. For the current study, we took advantage of a recently developed DNA-enrichment method to produce four complete genomes and two draft genomes of *Wolbachia* from filarial nematodes. Two complete genomes, *wCtub* and *wDcau*, are the smallest *Wolbachia* genomes sequenced to date (863988 bp and 863427 bp, respectively), as well as the first genomes representing supergroup J. These genomes confirm the validity of this supergroup, a controversial clade due to weaknesses of the multilocus sequence typing approach. We also produced the first draft *Wolbachia* genome from a supergroup F filarial nematode representative (*wMhie*), two genomes from supergroup D (*wLsig* and *wLbra*) and the complete genome of *wDimm* from supergroup C. Our new data confirm the paradigm of smaller *Wolbachia* genomes from filarial nematodes containing low levels of transposable elements and the absence of intact bacteriophage sequences, unlike many *Wolbachia* from arthropods, where both are more abundant. However, we observe differences among the *Wolbachia* genomes from filarial nematodes: no global co-evolutionary pattern, strong synteny between supergroup C and supergroup J *Wolbachia*, and more transposable elements observed in supergroup D *Wolbachia* compared to the other supergroups. Metabolic pathway analysis indicates several highly conserved pathways (haem and nucleotide biosynthesis, for example) as opposed to more variable pathways, such as vitamin B biosynthesis, which might be specific to certain host–symbiont associations. Overall, there appears to be no single *Wolbachia*–filarial nematode pattern of co-evolution or symbiotic relationship.

Received 07 August 2020; Accepted 14 November 2020; Published 09 December 2020

Author affiliations: ¹Molecular Parasitology Group, New England Biolabs, Ipswich, MA, USA; ²Instituto de Zoología y Ecología Tropical, Universidad Central de Venezuela, Caracas, Venezuela; ³Ediciones La Fauna KPT SL, Madrid, Spain; ⁴Neotropical Parasitology Research Network - NEOPARNET, Asociación Peruana de Helmintología e Invertebrados Afines – APHIA, Peru; ⁵Epidemiology, Parasites and Vectors, ARC-Onderstepoort Veterinary Institute, Onderstepoort 0110, South Africa; ⁶Unité Molécules de Communication et Adaptation des Microorganismes (MCAM, UMR7245), Muséum National d'Histoire Naturelle, CNRS, Paris, France; ⁷Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK.

*Correspondence: Barton E. Slatko, barton.slatko@gmail.com

Keywords: filarial nematodes; genomics; symbiosis; target enrichment; *Wolbachia*.

Abbreviations: ANI, average nucleotide identity; CCS, circular consensus sequence; CDS, coding sequence; CI, cytoplasmic incompatibility; dDDH, digital DNA–DNA hybridization; DDH, DNA–DNA hybridization; HR, homologous recombination; IS, insertion sequence; LEFT-SEQ, large enriched fragment targeted sequencing; ML, maximum likelihood; MLST, multilocus sequence typing; MNHN, National Museum of Natural History; NCBI, National Center for Biotechnology Information; ORI, origin of replication; PCo, principal coordinates.

†Present address: School of Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, AZ, USA.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Eleven supplementary tables and three supplementary files are available with the online version of this article.

000487 © 2020 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

DATA SUMMARY

Data generated are available in the National Center for Biotechnology Information (NCBI) databases: BioProject PRJNA593581; BioSample SAMN13482485 for wLsig, *Wolbachia* endosymbiont of *Litomosoides sigmodontis* (genome: CP046577); BioSample SAMN15190311 for the nematode host *Litomosoides sigmodontis* (genome: JABVXW000000000); BioSample SAMN13482488 for wDimm, *Wolbachia* endosymbiont of *Dirofilaria (Dirofilaria) immitis* (genome: CP046578); BioSample SAMN15190314 for the nematode host *Dirofilaria (Dirofilaria) immitis* (genome: JABVXT000000000); BioSample SAMN13482046 for wCtub, *Wolbachia* endosymbiont of *Cruorifilaria tubero-cauda* (genome: CP046579); BioSample SAMN15190313 for the nematode host *Cruorifilaria tubero-cauda* (genome: JABVXU000000000); BioSample SAMN13482057 for wDcau, *Wolbachia* endosymbiont of *Dipetalonema caudispina* (genome: CP046580); BioSample SAMN15190312 for the nematode host *Dipetalonema caudispina* (genome: JABVXV000000000); BioSample SAMN13482459 for wLbra, *Wolbachia* endosymbiont of *Litomosoides brasiliensis* (genome: WQMO000000000); BioSample SAMN15190311 for the nematode host *Litomosoides brasiliensis* (genome: JABVXW000000000); BioSample SAMN13482487 for wMhie, *Wolbachia* endosymbiont of *Madathamugadia hiepei* (genome: WQMP000000000); BioSample SAMN15190315 for the nematode host *Madathamugadia hiepei* (genome: JABVXS000000000). The raw data are available in the NCBI Sequence Read Archive (SRA): SRR10903008 to SRR10903010; SRR10902913 to SRR10902914; SRR10900508 to SRR10900511; SRR10898805 to SRR10898806.

INTRODUCTION

The endosymbiotic alpha-proteobacterium *Wolbachia* represents a striking model for studies of symbioses. These bacteria have been detected in a large proportion of arthropods, where they are considered one of the most widespread symbionts [1, 2], and in only two divergent families of parasitic nematodes (filarial nematodes in vertebrates and pratylenchid nematodes feeding on plants) [3, 4]. The nature of the relationships with their hosts is particularly fascinating. In arthropods, some *Wolbachia* are reproductive parasites inducing different phenotypes such as cytoplasmic incompatibility (CI), male-killing, parthenogenesis or feminization of genetic males [5–8]. Others *Wolbachia* are nutritional mutualists, as in the case of bedbug symbionts [9], while in the case of the filarial nematodes, their *Wolbachia* are obligate mutualists [10]. Numerous *Wolbachia* genomes have been subjected to genomic analyses to determine the nature of the symbiosis [11–15]. Some candidate genes have been identified as being involved in CI [16, 17], male-killing [18], feminization [19] and nutritional supplementation [20, 21]. The CI phenotype is being exploited as a tool for human disease prevention of mosquito-borne diseases as it is able to suppress pathogens, notably RNA arboviruses (such as dengue, chikungunya, Zika and yellow fever, and potentially other human pathogens,

Impact Statement

Wolbachia are endosymbiotic bacteria infecting a large range of arthropod species and two different families of nematodes, characterized by causing diverse phenotypes in their hosts, ranging from reproductive parasitism to mutualism. While available *Wolbachia* genomic data are increasing, they are not representative of the full *Wolbachia* diversity; indeed, 84% of *Wolbachia* genomes available from the National Center for Biotechnology Information database to date belong to the two main studied clades (supergroups A and B, exclusively composed of *Wolbachia* from arthropods). The present study presents the assembly and analysis of four complete genomes and two draft genomes of *Wolbachia* from filarial nematodes. Our genomic comparisons confirm the paradigm that smaller *Wolbachia* genomes from filarial nematodes contain low levels of transposable elements and the absence of intact bacteriophage sequences, unlike many *Wolbachia* from arthropods. However, data show disparities among the *Wolbachia* genomes from filarial nematodes: no single pattern of co-evolution, stronger synteny between some clades (supergroups C and supergroup J) and more transposable elements in another clade (supergroup D). Metabolic pathway analysis indicates both highly conserved and more variable pathways, such as vitamin B biosynthesis, which might be specific to certain host–symbiont associations. Overall, there appears to be no single *Wolbachia*–filarial nematode pattern of symbiotic relationship.

such as *Plasmodium*) [22–27]. With respect to the filarial nematodes, much of the research effort has been focused on drug screening or various treatments targeting *Wolbachia* to kill the parasitic species responsible for human diseases of lymphatic filariasis (elephantiasis) or onchocerciasis (river blindness), which are a major cause of global morbidity [13, 28–32]. Anti-filarial *Wolbachia* screening projects using genomic information [28, 33–37] and/or mass screening of chemicals, drugs and biomolecules from varied pre-existing, diversified or focused molecular libraries that select inhibitors of *Wolbachia* development and reproduction are underway [13, 30, 38–42]. Nonetheless, the mechanisms underpinning the obligate mutualism between *Wolbachia* and their filarial hosts remain largely unknown.

To date, four complete genomes of *Wolbachia* from filarial nematodes have been published: first, the symbiont of the human parasite *Brugia malayi*, wBm [33], followed by the symbiont of the bovine parasite *Onchocerca ochengi*, wOo [43], then the symbiont of the human parasite *Onchocerca volvulus*, wOv [44], and recently the symbiont of zoonotic parasite *Brugia pahangi*, wBp [45]. The hypothesis of potential provisioning of resources [such as haem, riboflavin, flavin adenine dinucleotide (FAD) or nucleotides] by *Wolbachia* to

their host nematodes has been suggested, beginning with the first comparative genomic analysis [33]. However, the analysis of the highly reduced *wOo* genome did not strongly support the hypothesis of provisioning of vitamins or cofactors by this strain (the riboflavin metabolism and FAD pathways are incomplete), and transcriptomic analysis suggested more of a role in energy production and modulation of the vertebrate immune response [43]. Alternatively, the relationship between filarial nematodes and *Wolbachia* may represent a ‘genetic addiction’ rather than genuine mutualism [46]. *Wolbachia* genomes from filarial nematodes as compared to *Wolbachia* genomes from arthropods have smaller sizes [between 957990 bp for *wOo* to 1080084 bp for *wBm* versus 1250060 bp for *wMel* (from *Drosophila melanogaster*), 1267782 bp for *wCle* (from *Cimex lectularius*), 1587994 bp for *wPip* (from *Culex pipiens*) and 1801626 bp for *wFol* (from *Folsomia candida*)]. In addition, the presence of fewer transposable elements [such as insertion sequence (IS) elements and group II intron-associated genes], prophage-related genes and repeat-motif proteins (such as ankyrin domains) has been described [43, 47].

The notion of *Wolbachia* species remains under debate within the scientific community [48–51]. It is commonly accepted to describe the various *Wolbachia* strains as belonging to different phylogenetic lineages as ‘supergroups’ (currently A–F, H–Q and S). In the past, two *Wolbachia* supergroups G and R had been demonstrated as invalid based on further phylogenetic analyses [52, 53]. The first appearance of the ‘supergroups’ designation dates to 1998 [54], but the concept was popularized later by Lo *et al.* [55]. Most of the molecular characterizations of *Wolbachia* strains have been based on either single gene or multi-locus phylogenies [53, 55–65]. The supergroups A, B, E, H, I, K, M, N, O, P, Q and S are exclusively composed of symbionts of arthropods [55, 57, 59, 63, 66–70]. In contrast, supergroups C, D and J are restricted to filarial nematodes [4, 58, 61, 71], whereas supergroup L is found only in plant-parasitic nematodes [3, 72]. Supergroup F is, so far, the only known clade comprising symbionts of filarial nematodes as well as arthropods [55, 56, 73]. Initially, the delimitation of these supergroups was defined arbitrarily by a threshold of 2.5% divergence of the *Wolbachia* surface protein gene (*wsp*) [54]. However, after it was demonstrated that *wsp* could recombine between *Wolbachia* strains [74], a multilocus sequence typing (MLST) approach for *Wolbachia*

was proposed [60]. These typing methods were developed based almost exclusively on analyses of supergroup A and B *Wolbachia* [60], as they constituted the majority of genomes sequenced at that time. Recently, there has been an effort to revisit the MLST paradigm [75] and attempts to classify *Wolbachia* based on genomics [50]. However, increased genomic information is needed to appraise the phylogenetic diversity of *Wolbachia* representatives from filarial nematodes. The presently available *Wolbachia* genomic information is not fully representative of *Wolbachia* diversity.

Recently, a method based on biotinylated probes was developed to capture large fragments of *Wolbachia* DNA for sequencing, using PacBio technology (large enriched fragment targeted sequencing – LEFT-SEQ) [76] adapted from previous capture methods using Illumina technology [77, 78]. We used this enrichment method to produce draft or complete genomes of *Wolbachia* from a diversity of filarial nematode species: *Cruorifilaria tubero cauda*, a parasite of the capybara, a cavy rodent; *Dipetalonema caudispina*, a parasite of spider monkeys; *Litomosoides brasiliensis*, a parasite of bats; *Litomosoides sigmodontis*, a parasite of cricetid rodents; *Dirofilaria (Dirofilaria) immitis*, a parasite of canines; and *Madathamugadia hiepei*, a parasite of geckos. These species had been previously characterized as positive for *Wolbachia* infections (two supergroup J, two supergroup D, one supergroup C and one supergroup F) [58]. In the present study, we took advantage of this newly explored diversity to draw a more comprehensive picture of symbiosis between *Wolbachia* and their filarial nematode hosts.

METHODS

Materials

Eight specimens belonging to six filarial nematode species were studied (Table 1): *Cruorifilaria tubero cauda* (two samples), *Dipetalonema caudispina* (two samples), *Litomosoides brasiliensis*, *Litomosoides sigmodontis*, *Dirofilaria (Dirofilaria) immitis* and *Madathamugadia hiepei*.

The DNA samples of *Cruorifilaria tubero cauda*, *Dipetalonema caudispina*, *Litomosoides brasiliensis* and *Madathamugadia hiepei* were provided by the National Museum of Natural History (MNHN), Paris, France. These DNA samples had been extracted from adult worms for a previous study [79].

Table 1. Information on the species studied

| Species | Host | Specimen | Collection locality |
|--|----------------------------------|---------------|-----------------------------|
| <i>Cruorifilaria tubero cauda</i> | <i>Hydrochoerus hydrochaeris</i> | 55YT/56YT | Venezuela |
| <i>Dipetalonema caudispina</i> | <i>Ateles paniscus</i> | 362YU2/362YU3 | Guyana |
| <i>Dirofilaria (Dirofilaria) immitis</i> | <i>Canis familiaris</i> | – | FR3 strain (GA, USA) |
| <i>Litomosoides brasiliensis</i> | <i>Carollia perspicillata</i> | 37PF | Peru |
| <i>Litomosoides sigmodontis</i> | <i>Meriones unguiculatus</i> | – | MNHN strain (Paris, France) |
| <i>Madathamugadia hiepei</i> | <i>Chondrodactylus turneri</i> | 81YU | South Africa |

The specimens had been donated to the MNHN by hunters or veterinarians and all procedures were conducted in compliance with the rules and regulations of the respective national ethical bodies [79]. The *Dirofilaria (Dirofilaria) immitis* specimen was provided by the NIAID/NIH Filariasis Research Reagent Resource Center (MTA University of Wisconsin-Oshkosh, Oshkosh, WI, USA; www.filariasiscenter.org), and the *Litomosoides sigmodontis* specimen was provided by the MNHN, where the experimental procedures were carried out in strict accordance with the European Union Directive 2010/63/UE and the relevant national legislation (ethical statement no. 13845). Supplementary file S1 (available with the online version of this article) lists the author(s) and year of parasite and host species collection. In accordance with the generally used nomenclature of *Wolbachia* strains, we have named these newly typed strains after their hosts: *wCtub* for the symbiont of *Cruorifilaria tubero cauda*; *wDcau* for the symbiont of *Dipetalonema caudispina*; *wLbra* for the symbiont of *Litomosoides brasiliensis*; *wLsig* for the symbiont of *Litomosoides sigmodontis*; *wDimm* for the symbiont of *Dirofilaria (Dirofilaria) immitis* and *wMhie* for the symbiont of *Madathamugadia hiepei*. In the cases of *wLsig* and *wDimm*, they have been referred to as *wLs* and *wDi*, respectively, in the prior literature [44, 80–82]. However, the lack of a consensus on the nomenclature of *Wolbachia* strains has already led to some confusion, as *wDi* can refer to *Wolbachia* from *Dirofilaria (Dirofilaria) immitis* [81] as well as *Wolbachia* from *Diaphorina citri* [83]. Therefore, to avoid confusion, we use the longer strain abbreviations in this paper.

The DNA of the *Madathamugadia hiepei* sample 81YU had been extracted previously and conserved at -20°C for 8 years [79]. The DNA of the *Dipetalonema caudispina* sample 362YU2 had been extracted in January 2015 and then conserved at -20°C for 4 years (unpublished data). A new

fragment of a specimen from the same lot (362YU3) was also obtained for new DNA extraction. DNA was extracted from all samples using the DNeasy kit (Qiagen), following the manufacturer's recommendations, including overnight incubation at 56°C with proteinase K.

Library preparations

According to the amount and quality of DNA of each sample, different library preparation protocols were utilized (Table 2). We used capture enrichment methods for either Illumina or PacBio sequencing based on the use of biotinylated probes to capture *Wolbachia* DNA (probes designed by Roche; NimbleGen) based on 25 complete or draft sequences as described by Lefoulon et al. [76]. The LEFT-SEQ method [76], developed for PacBio sequencing, was used for the freshly extracted DNA of *Dirofilaria (Dirofilaria) immitis*, *Litomosoides sigmodontis*, *Cruorifilaria tubero cauda* (55YT) and the 4-year-old extracted DNA of *Dipetalonema caudispina* (362YU2). We used $1\ \mu\text{g}$ DNA for each sample. Regarding the enriched libraries from *Dirofilaria (Dirofilaria) immitis* and *Litomosoides sigmodontis*, the last steps were modified, compared to the previously described protocol [76]: after the second PCR amplification of the enriched DNA, the libraries were prepared with the SMRTbell express template kit v2.0 (PacBio) prior to performing PacBio Sequel sequencing.

For *Madathamugadia hiepei* and *Litomosoides brasiliensis*, the DNA concentrations were either of low concentration or too fragmented to be used for the LEFT-SEQ protocol. For these, the enrichment method adapted for Illumina sequencing was a hybrid protocol between the method described by Geniez et al. [78] and that of Lefoulon et al. [76]. We followed the same procedure for the freshly extracted DNA of *Dipetalonema caudispina* (362YU2). DNA samples ($50\ \text{ng}$ for 81YU, $75\ \text{ng}$

Table 2. Sequencing data information

PacBio data are the number of CCSs produced with three full passes and a minimum predicted accuracy superior of 90%. Illumina data are the number of reads after filtering. Numbers in parentheses indicate the length of end-sequence protocol used.

| Species | Sample | PacBio LEFT-SEQ (CCSs) | Sequel PacBio (no capture) (CCSs) | Illumina (capture) | Illumina (no capture) |
|--|-------------|------------------------|-----------------------------------|---|-----------------------|
| <i>Cruorifilaria tubero cauda</i> | 55YT | 179618 | – | – | – |
| | 56YT | – | 124485 | – | 13205453 (2×100) |
| <i>Dipetalonema caudispina</i> | 362YU2 | 246679 | – | – | 25441996 (2×300) |
| | 362YU3 | – | 6446 | 5567787 (2×75); 4789888 (2×250) | – |
| <i>Dirofilaria (Dirofilaria) immitis</i> | FR3 strain | 155017 | 130552 | – | 118681906 (1×150) |
| <i>Litomosoides sigmodontis</i> | MNHN strain | 180870 | 157943 | – | 325671309 (1×150) |
| <i>Litomosoides brasiliensis</i> | 37PF | – | 8134 | 13247536 (2×75); 18143613 (2×250); 12523494 (2×150) | 14112979 (1×300) |
| <i>Madathamugadia hiepei</i> | 81YU | – | – | 1797778 (2×75); 1753231 (2×250); 1072731 (2×150) | 176201439 (1×150) |

for 37PF and 100 ng for 362YU3) were fragmented using the NEBNext Ultra II FS DNA kit (New England Biolabs) at 37°C for 20 min (resulting in DNA fragments with a mean size of 350 bp). The sheared DNA samples were independently ligated to SeqCap barcoded adaptors (NimbleGen; Roche), to enable processing of multiple samples simultaneously. The ligated DNAs were amplified by PCR and hybridized to the biotinylated probes, according to the SeqCap EZ HyperCap protocol (Roche NimbleGen user's guide v1.0). For each sample, a library without the enrichment method was also processed using a NEBNext Ultra II FS DNA library prep kit, following the manufacturer's recommendations (New England Biolabs). The library preparation for *Dipetalonema caudispina* sample 362YU2 was performed in 2016 at the University of Liverpool. Supplementary libraries without enrichment for PacBio sequencing were also processed for *Cruorifilaria tuberocauda*, *Dipetalonema caudispina* and *Litomosoides brasiliensis* using the SMRTbell express template prep kit v2.0, following the manufacturer's recommendations (PacBio). The data produced are summarized in Table 2.

De novo assembly pipeline

The bioinformatics pipeline was slightly different for the six samples, because of the variations in sequence protocols (as described above). However, the pre-processing of the reads was similar for both Illumina and PacBio data. The Illumina reads were filtered using the wrapper Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). For PacBio, circular consensus sequences (CCSs) were generated using the SMRT pipe RS_ReadsOfInsert protocol (PacBio) with a minimum of three full passes and minimum predicted accuracy greater than 90%. The adapter and potential chimeric reads were removed using seqtk (github.com/lh3/seqtk) as described by Lefoulon *et al.* [76] (analyses were performed with an in-house shell script). When sufficient PacBio reads were obtained, a *de novo* long-read assembly was done using Canu [84] according to Lefoulon *et al.* [76]. Otherwise, a first hybrid *de novo* assembly was done using Spades [85]. The contigs belonging to *Wolbachia* were detected by nucleotide similarity using BLASTN (similarity greater than 80%, bitscore greater than 50) [86] and isolated. The remaining contigs were manually curated to eliminate potential non-*Wolbachia* sequence contaminations. To improve the complete assembly of the *Wolbachia* genomes, a selection of reads mapping to the first *Wolbachia* draft genome was produced. The Illumina reads were merged with PEAR [87] (in the case of end-paired reads) and mapped against this contig selection using Bowtie2 [88]. The PacBio reads were mapped against this contig selection using ngmlr (with the PacBio pre-set settings) [89]. A second hybrid *de novo* assembly was then performed with this new selection of reads using Unicycler [90]. A second selection of *Wolbachia* contigs using BLASTN was then performed and manually curated to eliminate potential contaminations. Assembly statistics were calculated using QUAST [91]. PCR primers were designed to confirm the sites of circularization of the single contigs when applicable (Supplementary file S2).

In addition, contigs representative of filarial nematode genomes were isolated from the first *de novo* assembly. Mapped reads were selected, as described above, and *de novo* assemblies of the host genomes were produced using Unicycler [90].

The completeness of the draft genomes was studied using BUSCO v3, which analyses the gene content compared to a selection of near-universal single-copy orthologous genes. This analysis was based on 221 genes common among proteobacteria for the draft genomes of *Wolbachia* (proteobacteria_odb9), while it was based on 982 genes common among nematodes for the host draft genomes (nematoda_odb9).

Comparative genomic analyses and annotation

We used different comparative analyses between the produced draft genomes and a set of eight available complete genomes and seven draft genomes of *Wolbachia* (Table S1): *wMel*, *Wolbachia* from *Drosophila melanogaster* (NC_002978), *wCau*, *Wolbachia* from *Carposina sasakii* (CP041215), and *wNfla*, *Wolbachia* from *Nomada flava* (LYUW00000000) for supergroup A; *wPip*, *Wolbachia* from *Culex quinquefasciatus* (NC_010981), *wTpre*, *Wolbachia* from *Trichogramma pretiosum* (NZ_CM003641), *wLug*, *Wolbachia* from *Nilaparvata lugens*, and *wstri* (MUIY01000000), *Wolbachia* from *Laodelphax striatella* (LRUH01000000) for supergroup B; *wVulC*, *Wolbachia* from *Armadillidium vulgare* (ALWU00000000), closely related to the supergroup B; *wPpe*, *Wolbachia* from *Pratylenchus penetrans* for supergroup L (NZ_MJMG01000000); *wCle*, *Wolbachia* from *Cimex lectularius* for supergroup F (NZ_AP013028); *wFol*, *Wolbachia* from *Folsomia candida* for supergroup E (NZ_CP015510); *wBm*, *Wolbachia* from *B. malayi* (NC_006833), *wBp*, *Wolbachia* from *B. pahangi* (NZ_CP050521), and *wWb*, *Wolbachia* from *Wuchereria bancrofti* (NJBR02000000), for supergroup D; *wOv* *Wolbachia* from *O. volvulus* (NZ_HG810405), and *wOo*, *Wolbachia* from *O. ochengi* (NC_018267) for supergroup C; *wCfeT* (NZ_CP051156.1) and *wCfeJ* (NZ_CP051157.1) both *Wolbachia* from *Ctenocephalides felis* (not described as belonging to any supergroup) [92].

We calculated the average nucleotide identity (ANI) between the different *Wolbachia* genomes using ANI Calculator [93] and an in-silico genome-to-genome comparison was done to calculate a digital DNA–DNA hybridization (dDDH) using GGDC [94]. The calculation of dDDH allows analysis of species delineation as an alternative to the wet-lab DNA–DNA hybridization (DDH) used for current taxonomic techniques. GGDC uses a genome BLAST distance phylogeny approach to calculate the probability that an intergenomic distance yielded a DDH larger than 70%, representing a novel species-delineation threshold [94]. We used formula two to calculate the dDDH, because it is more robust using incomplete draft genomes [95].

The *Wolbachia* genomes were analysed using the RAST pipeline [96]. In order to compare the nature of these genomes using the RAST pipeline, we identified the percentage of coding sequences (CDSs) (calculated by a ratio between total

base pairs of CDSs and total genome sequence base pairs), the presence of mobile elements, the group II intron-associated genes, the phage-like genes and the ankyrin-repeat protein genes. The presence of potential ISs was detected using ISSaga [97] (degraded sequences were not manually curated) and the presence of prophage regions was detected using PHASTER [98]. The *Wolbachia* genomes were annotated using Prokka [99]. We also examined the correlation between the size of the genome and the previously described genetic characteristics using the Spearman's rank correlation or Pearson rank correlation tests (if applicable after Shapiro–Wilk test) in the R environment [100]. KEGG orthology (KO) assignments were generated using KASS (KEGG Automatic Annotation Server) [101]. KASS assigned orthologous genes by a BLAST comparison against the KEGG genes database using the BBH (bi-directional best hit) method. The same assignment analysis was performed for the newly produced genomes and the set of 15 *Wolbachia* genomes from the National Center for Biotechnology Information (NCBI) database. The assigned KO were ordered in 160 different KEGG pathways (Table S2). Several pathways that showed differences in the number of assigned genes between *Wolbachia* genomes were selected. A list of genes assigned to these pathways was compiled to study potential losses/acquisitions of these genes across the various *Wolbachia* (Table S3).

Phylogenomic analyses

Single-copy orthologue genes were identified from a selection of *Wolbachia* genomes using Orthofinder (version 2.2.6) [102]. Three phylogenomic studies were performed: the first included only 11 *Wolbachia* genomes from filarial nematodes; the second included only 25 complete genomes, and the third included 49 complete or draft genomes (Table S1). The supermatrix of orthologue sequence alignments was generated by Orthofinder (implemented as functionality). The poorly aligned positions of the orthologous gene alignments produced were eliminated using Gblocks (version 0.91b) [103]. The phylogenetic analyses were performed with maximum-likelihood (ML) inference using IQ-TREE (version 1.5.5) [104]. The most appropriate model of evolution was evaluated by ModelFinder (implemented as functionality of IQ-TREE) [104]. The robustness of each node was evaluated by a bootstrap test (1000 replicates). The phylogenetic trees were edited in FigTree (<https://github.com/rambaut/figtree/>) and Inkscape (<https://inkscape.org/>). To study the evolution of the filarial hosts infected with *Wolbachia*, the same workflow was applied to the amino-acid files previously produced by the BUSCO analysis (Augustus implemented as functionality) based on the set of 982 orthologue genes common among nematodes (nematoda_odb9) (version 3.0.2) (Table S4). The six filarial nematode draft genomes produced were analysed with five draft genomes available in the database (AAQA00000000 for *B. malayi*; JRWH00000000 for *B. pahangi*; CAWC01000000 for *O. ochengi*, CBVM00000000 for *O. volvulus*, LAQH01000000 for *Wuchereria bancrofti*).

Synteny and co-evolutionary analyses

The potential positions of the origin of replication (ORI) were identified based on the ORI position in the *wMel* and *wBm* genomes according to Ioannidis *et al.* [105] for the complete genomes generated for *wDimm*, *wLsig*, *wDcau* and *wCtub*, as well as available complete genomes of *Wolbachia* from supergroup C (*wOo* and *wOv*), supergroup D (*wBp*) and supergroup F (*wCle*), and *wCfeJ*. The genome sequences were reorganized to start at the potential ORI position to study the genome rearrangement. Then, a pairwise genome alignment of these genomes was produced and plotted using MUMmer v3 [106].

Two global-fit methods were used to study the cophylogenetic pattern between filarial nematodes and their *Wolbachia* symbionts: PACo application [107] and ParaFit function [108] both in the R environment [100]. For these analyses, we independently produced two ML phylogenies: the phylogenetic tree of the 11 *Wolbachia* from filarial nematodes and the phylogenetic tree of the 11 filarial nematodes as described above. The global-fit method estimates the congruence between two phylogenetic trees changing the ML phylogenies into matrices of pairwise patristic distance, themselves transformed into matrices of principal coordinates (PCo). Then, PACo analysis transformed the symbiont PCo using least-squares superimposition (Procrustes analysis) to minimize the differences with the filarial PCo. The global fit was plotted in an ordination graph.

The congruence of the phylogenies was calculated by the residual sum of squares value (m^2_{xy}) of the Procrustean fit calculation. Subsequently, the square residual of each single association and its 95% confidence interval were estimated for each host–symbiont association and plotted in a bar chart [107]. A low residual value represented a strong congruence between symbiont and filarial host. In addition, the global fit was estimated using ParaFit function [108] in the R environment [100]. The ParaFit analysis tests the null hypothesis (H0), that the evolution of the two groups has been independent, by random permutations (1000000 permutations) of host–symbiont association [108]. This test is based on analysis of the matrix of patristic distances among the hosts and the symbionts as described above for PACo.

RESULTS

De novo assembly and completeness of draft genomes

We were able to produce complete circular assemblies for four of the genomes, *wCtub*, *wDcau*, *wLsig*, *wDimm*, as well as a 41-contig draft genome for *wLbra* and a 208-contig draft genome for *wMhie* (Table 3). The circularization of *wCtub*, *wDcau*, *wLsig* and *wDimm* was confirmed by PCR amplification of the sites of circularization of the single contigs (Supplementary file S2). The two supergroup J genomes, *wCtub* and *wDcau*, are the smallest observed among all sequenced *Wolbachia*, comprising 863988 and 863427 bp, respectively (Table 3). The genome of *wDimm*

Table 3. Draft genome information

| | Size (bp) | G+C content (mol%) | Contigs | N50 (bp) | BUSCO completeness (%) |
|-------|-----------|--------------------|---------|----------|------------------------|
| nDimm | 83711400 | 27.72 | 1294 | 126493 | 94.9 |
| wDimm | 920122 | 32.7 | 1 | 920122 | 76.0 |
| nLsig | 64161459 | 33.96 | 839 | 135680 | 94.1 |
| wLsig | 1045802 | 32.1 | 1 | 1045802 | 76.1 |
| nCtub | 75522022 | 30.29 | 1888 | 105487 | 95.6 |
| wCtub | 863988 | 32.3 | 1 | 863988 | 77.4 |
| nDcau | 81590899 | 30.44 | 2615 | 173032 | 97.7 |
| wDcau | 863427 | 28 | 1 | 863427 | 73.4 |
| nLbra | 65202511 | 31.74 | 1026 | 147890 | 93.4 |
| wLbra | 1046149 | 34.5 | 41 | 52864 | 76.5 |
| nMhie | 77701753 | 33.59 | 13165 | 17407 | 86.4 |
| wMhie | 1025329 | 36.1 | 208 | 6845 | 75.6 |

displayed a total size of 920122 bp and the total length of wLsig was 1045802 bp. The draft genome of wLbra had a total length of 1046149 bp, very close to the size observed for wLsig. Although the assembly remained fragmented, the draft genome of wMhie had a total length of 1025329 bp (Table 3). Varying success in producing complete genome sequences was attributed to DNA quantity and quality (Table 2); for example, the low quantity and quality of DNA obtained for *Madathamugadia hiepei* and *Litomosoides brasiliensis* limited sequencing. The *de novo* assembly was successful with production of a circularized genome in the case of wLsig based on 180 870 CCS PacBio reads, wDimm based on 155 017 CCS PacBio reads and wCtub based on 179 618 CCS PacBio reads. Of the sequenced CCS reads, 94.67% mapped to the draft genome for wLsig, 90.57% for wDimm, but only 35% for wCtub. However, for wDcau, the analysis of the sequenced 246 679 CCS PacBio reads using Canu did not produce an accurate draft genome (<100000 bp total length). In contrast, a hybrid *de novo* assembly, based on all the sequenced data, produced a draft genome containing one large *Wolbachia* contig, which was circularized using minimus2. Only 1.8% of the CCS reads produced with LEFT-SEQ (4466) mapped to this wDcau draft and the low efficiency was likely due to the 4-year-old extracted DNA that was used.

Among 221 single-copy orthologous genes conserved among proteobacteria (BUSCO database), 171 and 162 are present in the wCtub and wDcau complete genomes, respectively, suggesting 77.4 and 73.4% BUSCO completeness (Tables 3 and S4). The other two complete genomes, wLsig and wDimm, have 76.1 and 76% BUSCO completeness with 168 genes identified. The draft genome wLbra has 76.5% BUSCO completeness with 169 genes identified. The draft genome of wMhie has 75.6% BUSCO completeness with 167 genes identified. These levels of completeness are similar to most *Wolbachia* genomes from filarial nematodes. For

example, *Wolbachia* from *B. malayi*, wBm, has a higher level with 175 complete genes identified (79.2%) and *Wolbachia* from *O. ochengi*, wOo, has a lower level of completeness with 165 complete genes identified (74.7%) (Table S4). In general, *Wolbachia* genomes from arthropods present higher levels of BUSCO completeness [e.g. *Wolbachia* from *Drosophila melanogaster*, wMel, has 180 BUSCO genes (81.4%)]. The higher BUSCO completeness in these genomes could be because these genomes are less degraded than those of filarial *Wolbachia*.

Along with the assembly of the *Wolbachia* genome, draft genomes of the nematode hosts were produced (Table 3): a 1888-contig draft genome of *Cruorifilaria tubero cauda*, nCtub, of 75522022 bp total length; an 839-contig draft genome of *Litomosoides sigmodontis*, nLsig, of 64161459 bp total length; a 1026-contig draft genome of *Litomosoides brasiliensis*, nLbra, of 65202511 bp total length; a 1294-contig draft genome of *Dirofilaria (Dirofilaria) immitis*, nDimm, of 83711400 bp total length; a 2615-contig draft genome of *Dipetalonema caudispina*, nDcau, of 81590899 bp total length; and a 13165-contig draft genome of *Madathamugadia hiepei*, nMhie, of 77701753 bp total length. Among 982 single-copy orthologous genes conserved among nematodes (BUSCO database), the draft genome of *Dipetalonema caudispina* shows the highest level of completeness with 959 genes detected (97.7%). The draft genomes of *Cruorifilaria tubero cauda* and *Litomosoides brasiliensis* show similar results with 939 (95.6%) and 917 (93.4%) genes detected. The draft genome of *Madathamugadia hiepei* has the lowest level of completeness with 849 (86.4%) genes identified (Table S4). Draft genomes of nDimm and nLsig have previously been published with total lengths similar to these results, with 84888114 bp (ASM107739v1) and 64813410 bp (ASM90053727v1), respectively, but with lower N50 values (15147 and 45863, respectively) [109].

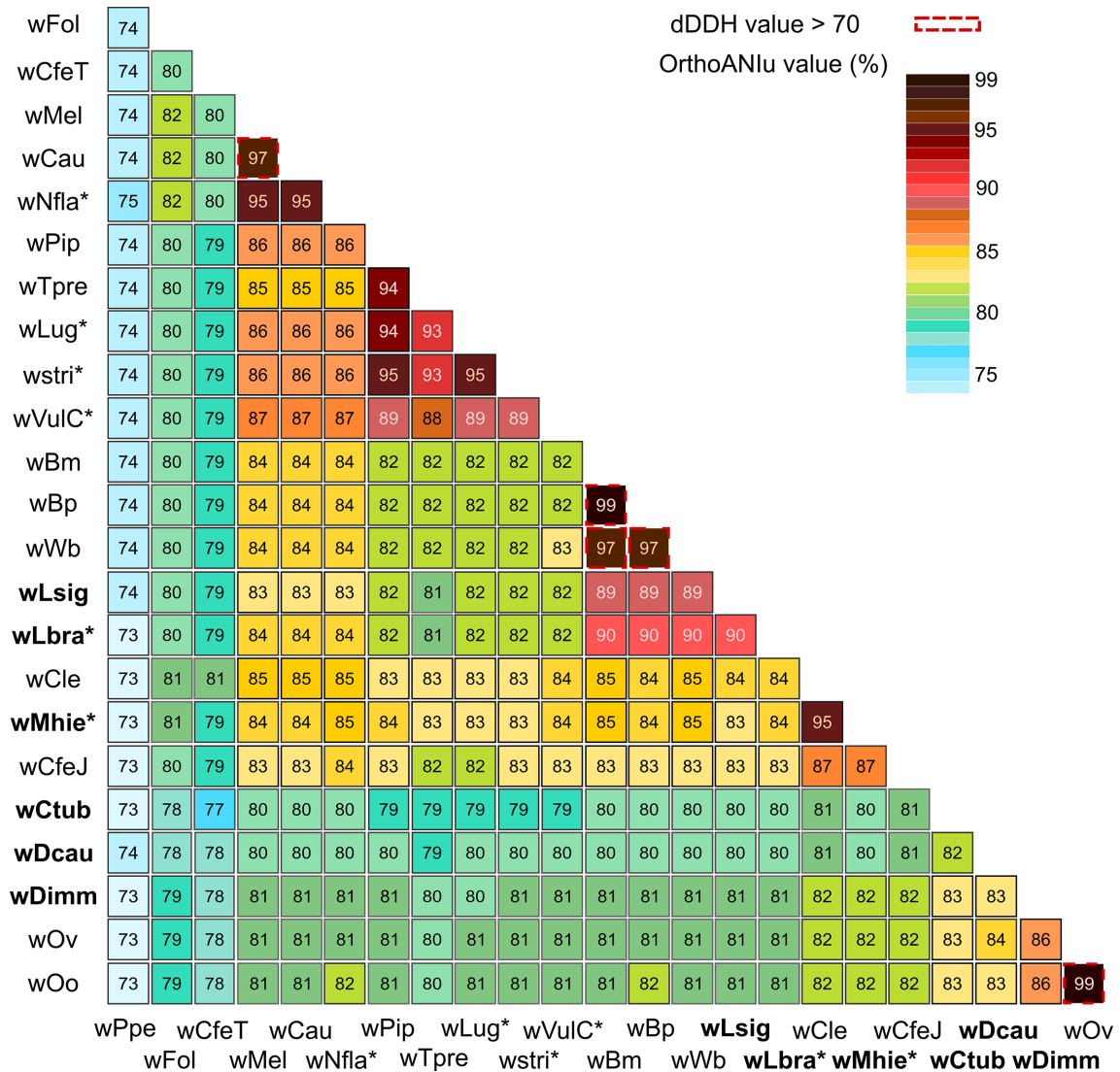


Fig. 1. Graphical representation of ANI and dDDH calculations for *Wolbachia* genomes. The ANI between 23 complete genomes of *Wolbachia* was evaluated using the ANI Calculator and the probability of a dDDH greater than 70 using ggdc. Asterisks represent draft genomes and the genome sequences produced in this study are indicated in bold.

ANI and dDDH

The ANI calculation indicates that *wCtub* and *wDcau* are divergent from other *Wolbachia*. For both, the most similar genome is *wOv* with 83 and 84 % identity, respectively (Fig. 1). The draft genome *wLbra* shows a stronger similarity of 90% with the representatives of supergroup D: *wBm*, *wBp*, *wWb* and *wLsig*. The draft genome *wMhie* is most similar to *wCle* from supergroup F with 95% identity (Fig. 1). Typically, strains representative of the same supergroup share strong identity: 99% for *wOo* and *wOv* (from supergroup C), 99% for *wBm* and *wBp* (from supergroup D), 97% for *wBm* or *wBp* and *wWb* (from supergroup D), 95% for *wPip* and *wstri* (from supergroup B) and 97% for *wMel* and *wCau* (from supergroup A). A dDDH [94] metric higher than 70 indicates that the two strains might belong to the same species (see Methods).

The present *in silico* genome-to-genome comparison shows only five cases that might be considered as similar strains of *Wolbachia*: *wCau* and *wMel*; *wBm* and *wBp*; *wBm* and *wWb*; *wBp* and *wWb*; and *wOo* and *wOv* (Fig. 1). These proximities have been suggested elsewhere [50]. Both ANI and dDDH analyses suggest that the four newly sequenced *Wolbachia* genomes are divergent from published *Wolbachia* genomes.

Phylogenomic analyses

A total of 367 single-copy orthologous genes were identified from among the 25 available complete *Wolbachia* genomes. The newly sequenced *wCtub*, *wDcau*, *wLsig* and *wDimm* genomes were included in the ML phylogenetic analyses based on these 367 orthologous genes (Fig. 2a). This phylogenetic analysis confirms that *wLsig* belongs to supergroup D

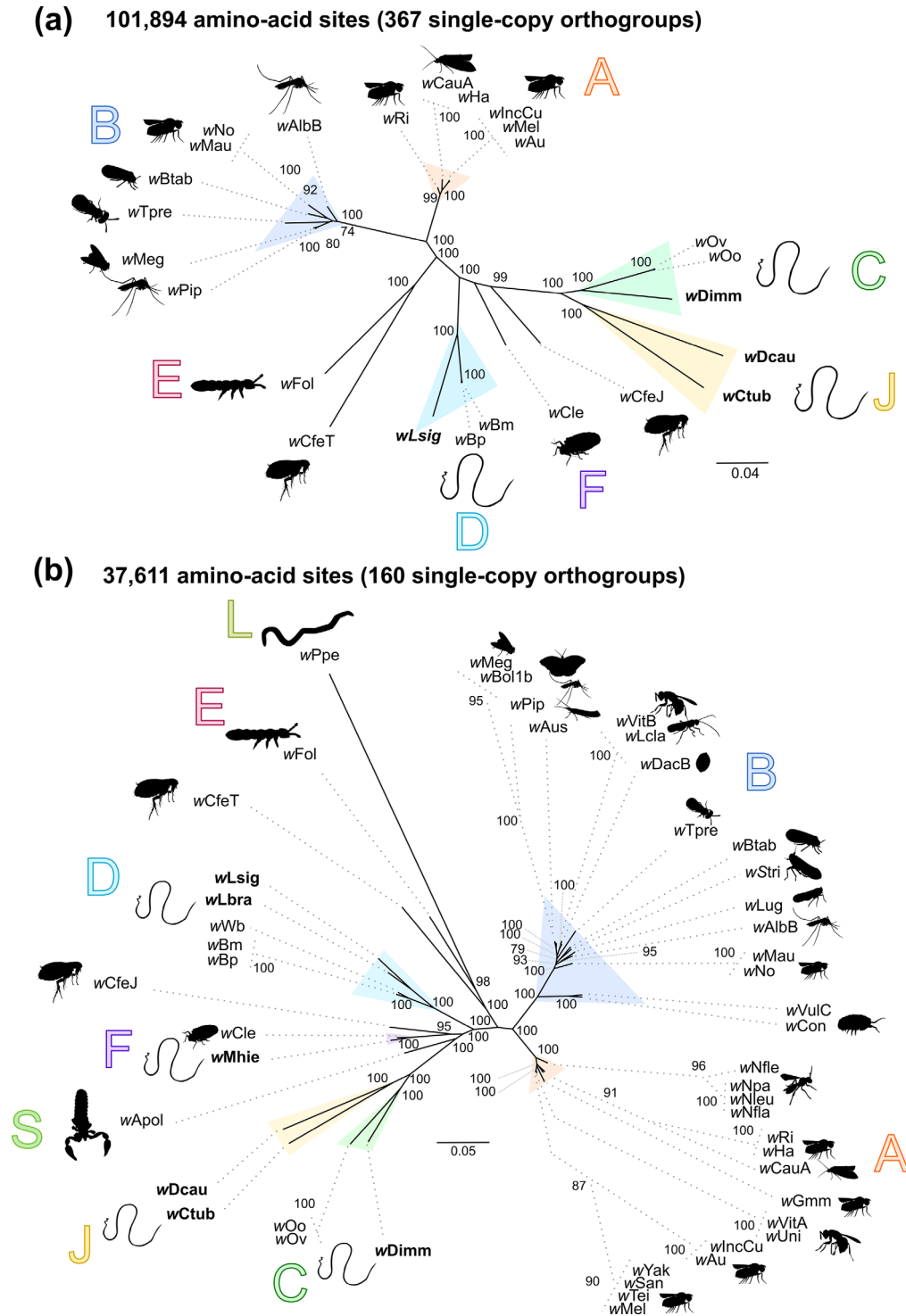


Fig. 2. Phylogenomic analyses of *Wolbachia*. The topologies were inferred using ML inference using IQ-TREE. Nodes are associated with bootstrap values based on 1000 replicates; only bootstrap values superior to 70 are indicated. The *Wolbachia* supergroups (A–L) are indicated. (a) Analysis based on concatenation of 365 single-copy orthogroups representing a 101894 amino acid matrix. The best-fit model calculated using ModelFinder according to the BIC index was JTT+F+I+G4. (b) Analysis based on concatenation of 160 single-copy orthogroups representing a 37611 amino acid matrix. The best-fit model calculated using ModelFinder according to the BIC index was JTT+F+I+G4. The scale bar indicates the distance in substitutions per nucleotide. The *Wolbachia* supergroups (A–L) are indicated and associated with different colours: orange for supergroup A, dark blue for B, light green for C, light blue for D, pink for E, purple for F, yellow for J, khaki green for L.

and *wDimm* belongs to supergroup C, as has been described elsewhere [4]. *wCtub* and *wDcau* were grouped in the same clade, a sister taxon of the supergroup C. *wCtub* and *wDcau* have been described as representatives of supergroup J [58]. The phylogenomic analysis presented here supports the hypothesis that *Wolbachia* supergroup J is a clade distinct from *Wolbachia* supergroup C, although the two clades are closely related. A total of 160 single-copy orthologous genes were identified among the 49 complete and draft *Wolbachia* genomes (Fig. 2b). The two phylogenomic analyses indicate the same topologies for the complete genomes *wCtub*, *wDcau*, *wLsig* and *wDimm*. In addition, the phylogenomic analysis based on 160 genes shows that *wLbra* is closely related to *wLsig*, as a representative of supergroup D, and *wMhie* is closely related to *wCle*, as a representative of supergroup F. The draft genome *wMhie* is the first representative of supergroup F infecting a filarial nematode and the phylogenomic analysis confirms the evolutionary history of *wLbra* and *wMhie*, as previously deduced from multi-locus phylogenies [58, 73].

The validity of supergroup J had been previously discussed; some studies using multi-locus phylogenies suggest that *Wolbachia* from *Dipetalonema gracile* (historically, the only known representative of supergroup J) belongs to supergroup C [57, 62, 110]. Interestingly, the *ftsZ* gene used in these MLST studies could not be detected in the *wDcau* and *wCtub* genomes. In addition, some multi-locus studies had observed PCR amplification of this gene to be unsuccessful within supergroup J *Wolbachia* [58, 111], while other studies included an *ftsZ* sequence from *Wolbachia* from *Dipetalonema gracile* [56]. To resolve this contradiction, we compared the 33 sequences of *Wolbachia* from *Dipetalonema gracile* available from the NCBI database and our complete *wDcau* genome, which should be closely related, using *NBLAST*. For 6 of these 33 sequences, from the *ftsZ* gene, it appears unlikely that they belong to *Wolbachia* from *Dipetalonema gracile*, only having between 72.37 and 89.91% identity with *wDcau* (while all other PCR sequences show 95.36–99.98% identity) (Table S5). Four of these six sequences are identical to genes of *Wolbachia* from *Drosophila* spp., one sequence is closely related to genes of *Wolbachia* from supergroup B [62] and one sequence is closely related to genes of *Wolbachia* from supergroup C (the *ftsZ* sequence) [56] (Table S5). Thus, our data suggest that the variable position of *Wolbachia* from *Dipetalonema gracile* in previous multi-locus phylogenies might be linked to contamination or errors of sequence submission.

Synteny conservation and co-evolutionary analysis

Strong conservation of synteny among supergroup C genomes and supergroup J genomes was observed (Fig. 3). It had been previously shown that the supergroup C genomes *wOo* and *wDimm* exhibit a low level of intra-genomic recombination [47]. Our results indicate a similar pattern of strong conservation of synteny among supergroup J genomes (*wDcau* and *wCtub*) and, more interestingly, between supergroup J genomes and *wDimm* in supergroup C (Fig. 3). This is in contrast to alignment of the complete genomic assemblies between the supergroup D genomes, which show more

rearrangement. Of further interest is the observation that a different level of rearrangement can be observed between *wBm* and *wLsig* or *wBp* and *wLsig*, even when *wBm* and *wBp* show less rearrangement between them. While *wBm* and *wBp* are characterized by a strong identity as described above (Fig. 1), similar to that observed between *wOo* and *wOv*, they show more rearrangement (Fig. 3).

The global-fit analyses do not show a global co-evolution pattern between filariae and their *Wolbachia* symbionts (PACo $m^2_{XY}=0.038$ with P value=1; ParaFitGlobal=0.0048 with P value=0.057; both 1×10^6 permutations). The superimposition plot indicates at least five groups of associations and shows strong inequality (Fig. 4a). The filarial nematodes *Dirofilaria* (*Dirofilaria immitis* and *Onchocerca* spp. with their symbionts (supergroup C) show lower squared residuals and consequently strong co-evolution. By contrast, *Madathamugadia hiepei* and its symbiont (supergroup F) show high squared residual and consequently a weak co-evolution (Fig. 4b). The global-fit analysis confirms two different groups of association for *Wolbachia* from supergroup D and their filarial nematode hosts: on one hand, *Brugia* and *Wuchereria* species and their symbionts, and on the other hand, *Litomosoides* species and their symbionts (Fig. 4a). The same trend is observed for *Wolbachia* from supergroup J; the filarial nematodes *Dipetalonema caudispina* and *Cruorifilaria tubero-cauda* and their symbionts present a higher squared residual than the residual sum of squares value (m^2_{XY}) suggesting a low congruence of the phylogenies (Fig. 4b). These results support the hypothesis of local patterns of co-evolution with multiple horizontal transmission events of *Wolbachia* among the filarial nematodes as part of the evolutionary history of this host–endosymbiont system, as previously described [58].

Comparative genomics

We observed a positive correlation between *Wolbachia* genome size and the percentage of CDSs. Indeed, *wDcau* and *wCtub* have the smallest genomes, and have a low percentage of CDSs (71.45 and 74.63%, respectively) (Fig. 5a). Similarly, a positive correlation was seen between *Wolbachia* genome size and transposable elements such as ISs, group II intron-associated genes and mobile elements (Fig. 5b, c, d). Interestingly, amongst the *Wolbachia* from filarial nematodes, supergroup C and supergroup J *Wolbachia* are all characterized by the absence or very low levels of transposable elements, unlike supergroup D *Wolbachia* and *wMhie* (supergroup F) (Fig. 6, Tables S6–S8). We also observed a positive correlation between *Wolbachia* genome size and the amount of insertion of phage DNA, as recently described (Fig. 5e, f) [112]. We studied phage DNA by two types of analyses: we used RAST annotation [96] to detect phage or phage-like genes and PHASTER [98] to detect prophage regions. None of the genomes of *Wolbachia* from filarial nematodes have significant prophage regions (Table S9) but supergroup D (*wBm*, *wBp* and *wWb*), as well as the supergroup F (*wMhie*) *Wolbachia* genomes, contain phage-like gene sequences inserted in their genomes. In the case of *wBm*, *wBp* and *wWb*, mainly phage major capsid protein and some

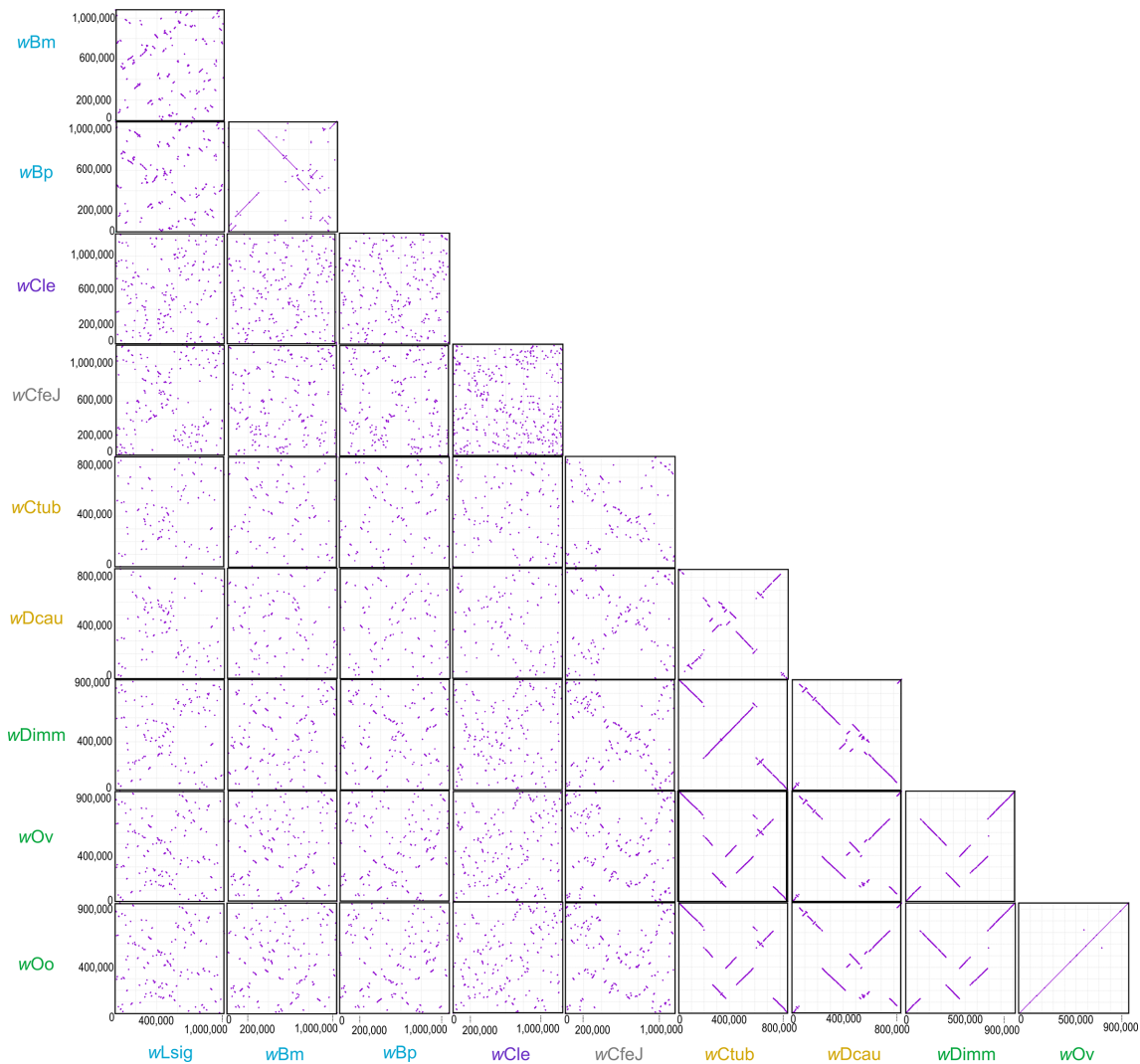


Fig. 3. Pairwise complete genome alignment for *Wolbachia* supergroups C, D, J and F and wCfeJ produced by MUMmer. The *Wolbachia* supergroups are indicated by different colours: light green for C, light blue for D, purple for F, yellow for J, and grey when no supergroup is assigned.

uncharacterized phage proteins were detected, representing 14 (total 2592bp), 8 (total 1197bp) and 4 regions (total 957bp), respectively (Table S10). The closely related wLsig and wLbra, also belonging to supergroup D, do not appear to have phage protein sequences. In the case of wMhie, one phage major capsid protein and eight other phage proteins were detected, representing 5187 bp (Table S10).

A positive correlation was also observed between *Wolbachia* genome size and ankyrin repeat proteins (Fig. 5g). It has been suggested that *Wolbachia* from filarial nematodes are characterized by a low level of ankyrin repeat genes, suspected to have evolved as result of their mutualistic lifestyle [113–115]. Most of the *Wolbachia* from filarial nematodes have 1 to 3 copies of ankyrin repeat genes, with the exception of wMhie (supergroup F) with 5 copies, and the supergroup D strains,

wBm, wBp and wWb, containing 14, 16 and 11 copies, respectively (Table S11).

Metabolic pathways

Using KAAS, we assigned genes from the 24 studied *Wolbachia* genomes (including the 7 draft genomes of wNfla, wLug, wstri, wVulC, wWb, wLbra and wMhie) to 160 different KEGG pathways (Table S2). Among these 160 KEGG pathways, 15 were selected based on strong variability among the genomes or because they had previously been suggested as being involved in symbiosis mechanisms [33, 43] (Fig. 7, Table 3).

In the context of a nutritional provisioning hypothesis, we observed variability among genomes in the vitamin B metabolism pathways (Fig. 7). The thiamine metabolism

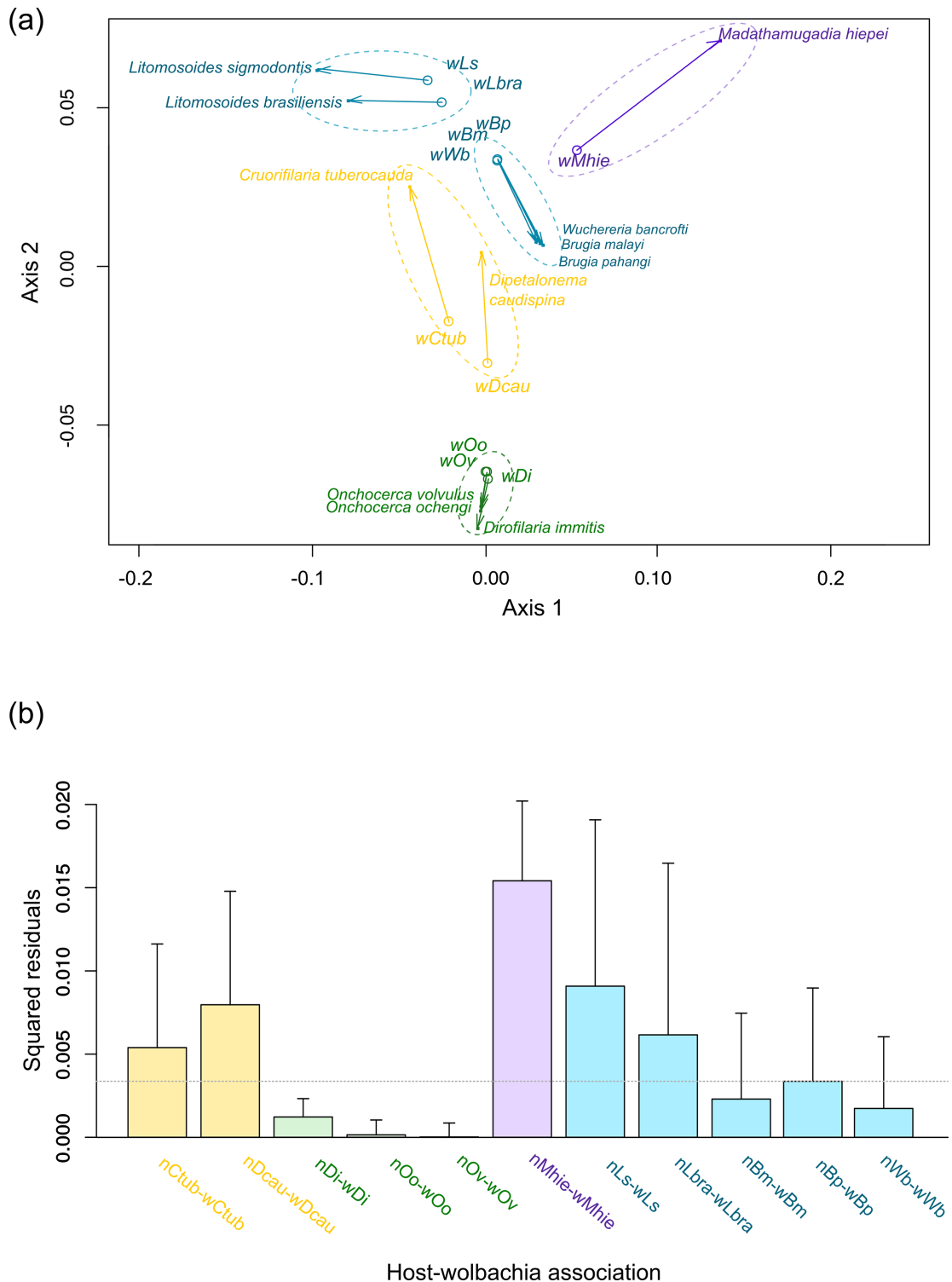


Fig. 4. Co-evolutionary analysis between filariae and *Wolbachia*. A PACo global-fit analysis of *Wolbachia* and their filarial host phylogenies was performed. (a) Representative plot of a Procrustes superimposition analysis, which minimizes differences between the two partners' principal correspondence coordinates of patristic distances. For each vector, the start point represents the configuration of *Wolbachia* and the arrowhead the configuration of filarial hosts. The vector length represents the global fit (residual sum of squares), which is inversely proportional to the topological congruence. (b) Contribution of each *Wolbachia*-filariae association to a general co-evolution. Each bar represents a Jackknifed squared residual and error bars represent upper 95% confidence intervals. The *Wolbachia* supergroups are indicated by different colours: light green for C, light blue for D, purple for F.

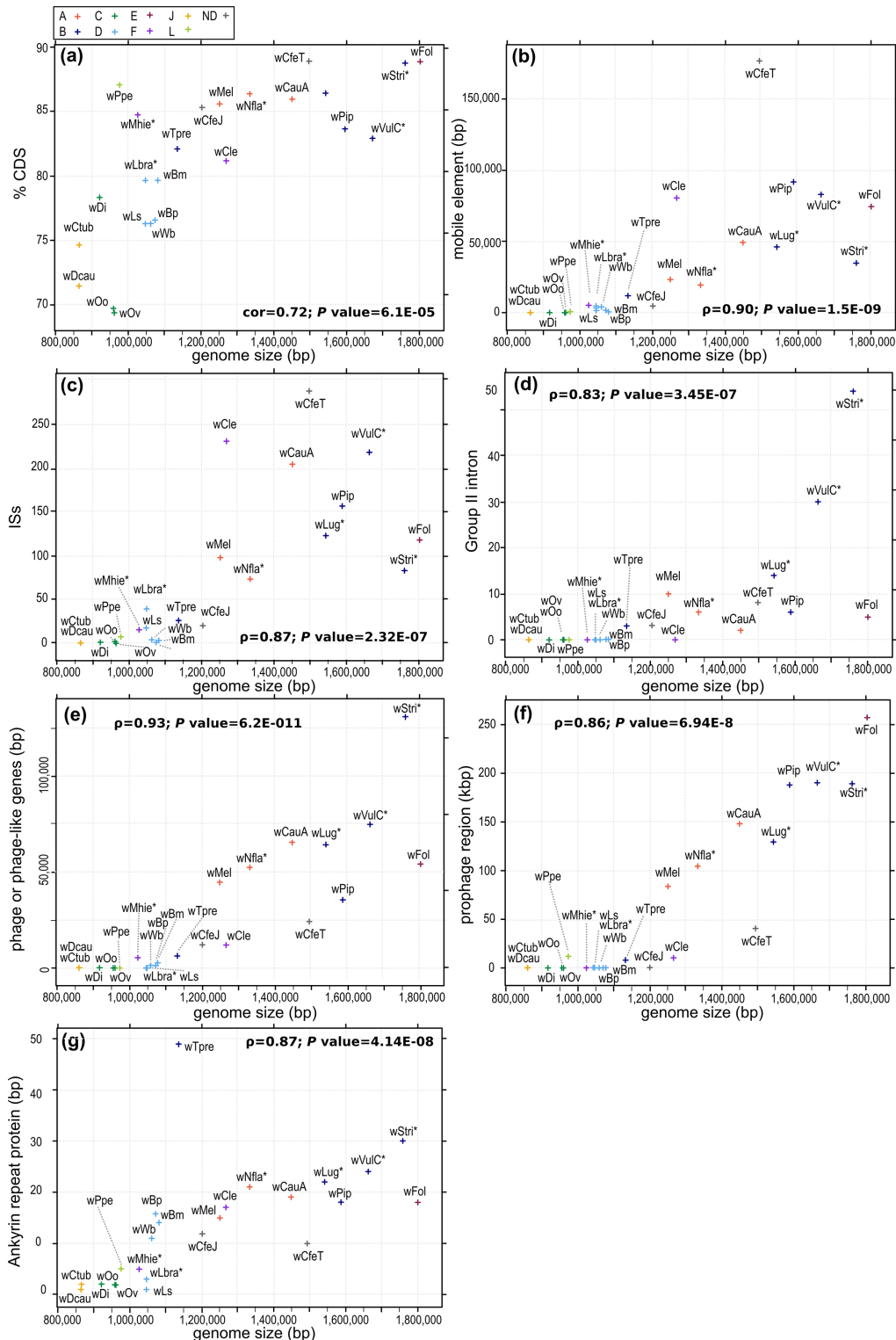


Fig. 5. Graphical representation of the relationship between the genome size of *Wolbachia* and different evolutionary factors: (a) percentage of CDSs, (b) regions identified as mobile elements, (c) regions identified as ISs, (d) regions identified as group II intron-associated genes, (e) regions identified as phage-like genes, (f) regions identified as potential prophages and (g) regions identified as ankyrin repeat regions. The *Wolbachia* supergroups (A–L) are indicated by different colours: orange for supergroup A, dark blue for B, light green for C, light blue for D, pink for E, purple for F, yellow for J, khaki green for L, and grey when the strain is not assigned to a supergroup.

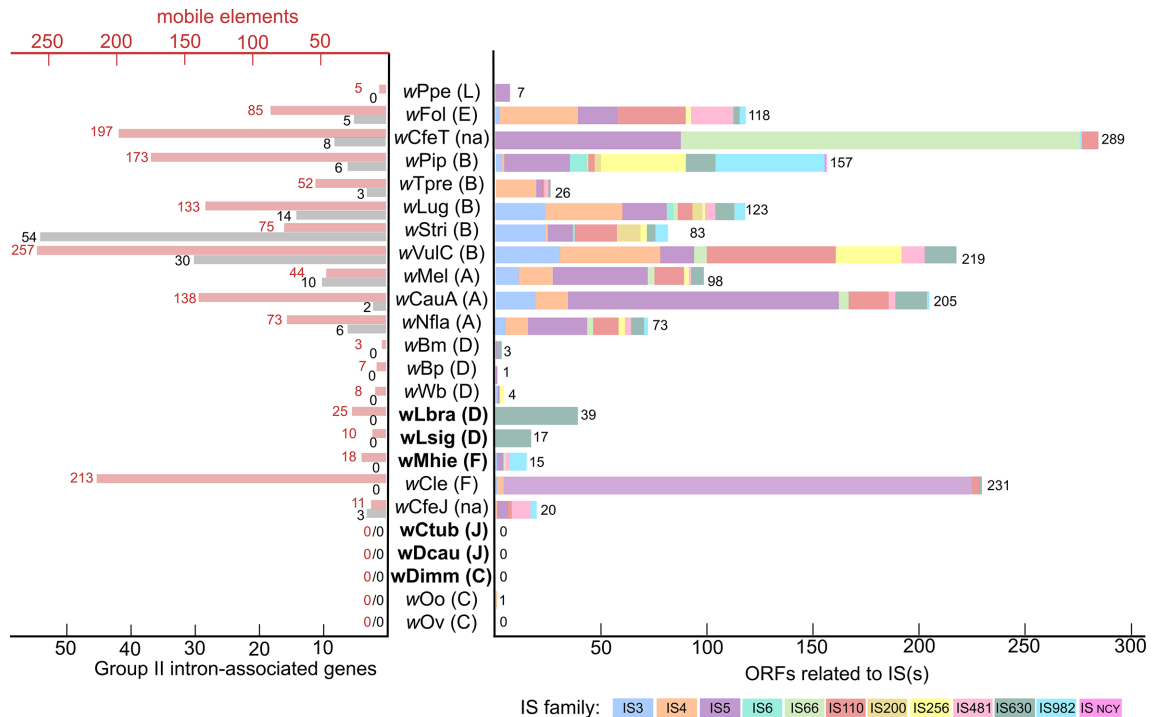


Fig. 6. Graph of IS elements, mobile elements and group II intron-associated genes identified in *Wolbachia* genomes. The *Wolbachia* supergroups are indicated in brackets: A to L and 'na' for *Wolbachia* without supergroup assignment. The graph on the left represents, in red, the number of mobile elements and, in black, the number of group II intron-associated genes detected in the studied genomes using RAST. The graph on the right represents the number of ORFs detected related to IS elements using ISSaga. For each of the detected ORFs related to an IS, the family of the IS is specified by a colour as indicated in the key below the graph. The number associated with each bar is the total number detected. Genomes produced in this study are indicated in bold.

(vitamin B1) pathway appears conserved, with the exception of the *tenA* gene, detected only in some *Wolbachia* genomes from arthropods: *wMel*, *wCauA*, *wNfla* (supergroup A); *wPip*, *wLug*, *wstri*, *wVulC* (supergroup B); *wFol* (supergroup E); *wCle* (supergroup F); *wCfeT* and *wCfeJ* (no supergroup designated). As previously reported by Darby *et al.* [43], the riboflavin metabolism (vitamin B2) is incomplete for both *wOo* (two genes not identified) and *wOv* (four genes not identified), but only *ribE* is missing for *wDimm*, another representative of supergroup C. Similarly, a single gene (*ribA*) is missing for *wDcau* and *wCtub* (supergroup J), as is the case for *wMhie* (supergroup F) and *wLbra* (supergroup D). The folate (vitamin B9) and pyridoxine (vitamin B6) metabolisms appear incomplete for representatives of supergroup D (B9, no *folA/B/C/KP* for all representatives; B6, absent for *wBp*, *wBm*, *wLsig*, but *pdxH* present for *wLsig* and *pdxH/pdxJ* present for *wWb*), but mainly conserved for other *Wolbachia* strains (except for *wPpe*, supergroup L, *wFol*, supergroup E, and *wCfeT*, in which only the *folC* gene is present in the folate pathway; and the absence of the *pdxJ* gene from *wOv* in the pyridoxine pathway) (Fig. 7). As described by other authors, we note that only some *Wolbachia* have a complete biotin metabolism pathway (vitamin B7): *wCle* (supergroup F) [20], *wNfla* (supergroup A) [116], *wstri* and *wLug* (supergroup B) [21], and *wCfeT* (no supergroup designated) [92]. In addition, we observed a complete biotin metabolism pathway in *wVulC*

(supergroup B). Interestingly, it has previously been suggested that supplementation of biotin by *Wolbachia* increases the fitness of insect hosts in the case of *wCle* from supergroup F [20], as well as *wstri* and *wLug* from supergroup B [21]. These genes could not be detected in the newly produced supergroup F genome, *wMhie*, from a filarial nematode host.

A further set of pathways previously considered of symbiotic interest, the *de novo* biosynthesis of purines and pyrimidines, has been identified in *Wolbachia* genomes, but was absent in other proteobacteria such as *Rickettsia* [33, 43]. The pyrimidine metabolism pathway was complete for most of the *Wolbachia* genomes analysed in the present study, with the exception of *wPpe* (supergroup L) (Fig. 7). The purine metabolism pathway was almost complete for the entire genome set as well, with the exception of the *purB* gene, which could not be identified in a large number of symbionts of filarial nematodes (*wLsig* and *wLbra*, supergroup D; all representatives of supergroups C and J; *wMhie*, supergroup F) and some symbionts of insects (*wNfla*, supergroup A; *wLug* and *wstri*, supergroup B; *wCle*, supergroup F; *wCfeJ* and *wCfeT*). The *purB* gene encoding the adenylosuccinate lyase protein is involved in the second step of the sub-pathway that synthesizes AMP from IMP.

Another important pathway, haem metabolism, suggested to be involved in symbiotic mechanisms by several genome

analyses, was complete in many of the current genomes. Only one gene, *bfr* (encoding bacterioferritin, a haem-storage protein), was not detected in *wBm* and *wWb* (supergroup D) or any representatives of supergroups C and J. The oxidative phosphorylation metabolism pathway also appears highly conserved, although the cytochrome *bd* ubiquinol oxidase genes (*cydA* and *cydB*) were detected only for *Wolbachia* belonging to supergroup A and *wCfeJ* (no supergroup assigned).

With regard to potential host interaction systems, the different secretion system pathways are very conserved (Fig. 7). However, the type II secretion system gene encoding the general secretion pathway protein D (*gspD*) was neither identified in *Wolbachia* belonging to supergroups C and J, nor in *wBm*, *wBp* and *wWb* (supergroup D). Similarly, the gene *secE* involved in the type Sec-SRP pathway was absent in *wNfla* (supergroup A), *wstri*, *wLug* (supergroup B), *wCfeT* and *wCfeJ* (no supergroups assigned).

A number of additional interesting variations among the studied *Wolbachia* genomes were noted, in particular for the cell cycle pathway, the homologous recombination (HR) pathway, the ATP binding cassette (ABC) transporter genes and glycerophospholipid metabolism (Fig. 7). Regarding the cell cycle pathway, representatives of supergroup J showed losses of most cell division proteins (only *ftsQ* was detected in *wCtub*), one gene of the two-component system (*pleD*), as well as the aspartyl protease family protein gene *perP*. The *ftsW* cell division protein gene was identified in only a few genomes: *wMel*, *wCauA*, *wNfla* (supergroup A); *wFol* (supergroup E); *wCle*, *wMhie* (supergroup F); *wLbra* (supergroup D); *wCfeT* and *wCfeJ*. Similarly, losses of numerous genes were detected in the HR metabolism pathway involved in repair of DNA damage. *Wolbachia* belonging to supergroup C, *wLsig* and *wLbra* (supergroup D), and *wDcau* (supergroup J) showed losses of numerous genes within this set (5–9 genes) (Fig. 7). We detected no recombination protein *rec* or Holliday junction DNA helicase *ruv* genes in the *wDcau*, *wOo* or *wOv* genomes. Another pronounced difference observed among the studied *Wolbachia* genomes was the presence of genes encoding ATP binding cassette (ABC) transporters. These membrane transporters appear largely depleted in the *Wolbachia* genome representatives of supergroups J and C, as well as in *wLsig* and *wLbra* (supergroup D) (Fig. 7). The haem exporter, phosphate transport system, lipoprotein-releasing system and zinc transport system appeared to be very conserved, unlike the biotin transport system, iron(III) transport system and phospholipid transport systems.

Regarding the glycerophospholipid metabolism, our results suggest that some genes are limited to a few genomes from arthropods. For example, the diacylglycerol kinase (ATP) gene (*dgkA*) was present in *wMel*, *wCauA*, *wNfla* (supergroup A), *wPip*, *wstri*, *wLug* (supergroup B), *wVulC* (closely related to supergroup B), *wFol* (supergroup E) and *wCfeT*, while the phospholipase D gene (*pld*) was only detected in *wVulC*, *wFol*, *wCfeJ*, *wLsig* and *wDimm* (Fig. 7).

DISCUSSION

The LEFT-SEQ method was applied to four invertebrate DNA samples, enabling us to produce four complete *Wolbachia* genomes: *wLsig*, *wDimm*, *wDcau* and *wCtub*. For *wLsig* and *wDimm*, draft genomes had previously been sequenced and analysed [47, 81, 82] but not submitted to the NCBI database. The complete genomes of *wDcau* and *wCtub* are, so far, the smallest *Wolbachia* genomes. Using the enrichment method associated with Illumina sequencing [78], two draft genomes were sequenced, a 41-contig *wLbra* draft genome and a 208-contig *wMhie* draft genome. Our data confirmed that *wLsig* and *wLbra* belong to supergroup D, *wDimm* resides in supergroup C, *wMhie* belongs to supergroup F, and *wDcau* and *wCtub* form a well-supported clade, supergroup J. Thus, *wDcau* and *wCtub* are now the first representative genomes of supergroup J, while *wMhie* constitutes the first representative genome of supergroup F from a filarial nematode. The ANI and dDDH index indicate that *wDcau*, *wCtub*, *wDimm*, *wLsig* and *wLbra* are clearly divergent from other studied *Wolbachia* genomes (all $\leq 90\%$ ANI and $< 70\%$ dDDH) (Fig. 1). Regarding *wMhie*, the ANI is 95% with *wCle*, suggesting a close genetic proximity, although the dDDH is equal to 58.12% (model-based confidence intervals: 76–82.5%), below the threshold of 70%. Our data suggest that these two *Wolbachia* are very similar, despite the fact that one infects a filarial nematode and the other infects bedbugs.

The analysis of supergroup J *Wolbachia* further highlights limitations of the current MLST system. In the past, the only representative identified from this supergroup was the symbiont of *Dipetalonema gracile*, a filarial nematode parasite of monkeys. This symbiont was first described as a deep branch within supergroup C [61], and subsequently as the divergent clade J [67, 111]. The latter phylogenetic position of this *Wolbachia* has been questioned by some authors and is often retained as belonging to supergroup C [57, 62, 110]. More recently, using a concatenation of seven genes and newly studied *Wolbachia*, Lefoulon et al. [58] demonstrated the validity of supergroup J, as distinct from supergroup C; a phylogenetic position confirmed by the present study (Fig. 2). Our analyses show that the *ftsZ* gene is not present (or is highly degenerate) in the *wDcau* and *wCtub* genomes, while previous *Wolbachia* phylogenies have been based on this marker. Our analyses (Table S5) suggest that the variable position of *Wolbachia* from *Dipetalonema gracile* in some phylogenetic analyses is linked to the fact that some database sequences likely do not belong to this strain.

The two complete genomes, *wDcau* and *wCtub*, are divergent from supergroup C *Wolbachia*. In addition, they are highly divergent from each other, with an ANI of 81%, despite the fact they form their own clade (Fig. 1). These divergences have been suggested by earlier multi-locus phylogenies with *Wolbachia* from *Cruorifilaria* and *Yatesia* species forming one subgroup, and *Wolbachia* from *Dipetalonema* spp. forming another subgroup within supergroup J [58]. Our data suggest that the use of the *ftsZ* gene for MLST studies is not appropriate for *Wolbachia* that are highly divergent. The fact that

the MLST system was designed on the basis of supergroups A and B *Wolbachia* [60], which have low genetic diversity [50], is a source of concern for its general use when studying divergent phylogenies. Moreover, the risk of erroneous data finding their way into databases (e.g. through contamination, mis-identification), combined with the fact that sequences used to build concatenated matrices very often do not originate from the same specimen, weakens multi-locus phylogenies, unless potential confounding factors are taken into consideration.

The symbiosis between *Wolbachia* and filarial nematodes was often considered and analysed as a uniform pattern of association, but our results reveal strong disparities. Indeed, the genomes of supergroup J present a strong synteny pattern, as was previously described for representatives of supergroup C, unlike those of supergroup D [47] (Fig. 3). We even observe a strong synteny pattern between *wDimm* (supergroup C) and *wCtub* or *wDcau* (supergroup J). Interestingly, the smaller *Wolbachia* genomes present a low number of genomic rearrangements, associated with the absence or low numbers of transposable elements (either ISs, mobile elements or group II introns) (Fig. 6). Our data support the paradigm that a major difference between *Wolbachia* from filarial nematodes and those from arthropods is a reduced genome containing fewer (or even zero) transposable elements, prophage-related genes or repeat-motif proteins (as ankyrin domains) [43]. Furthermore, our results highlight the distinction between supergroups C and J *Wolbachia* versus the supergroup D and F *Wolbachia*. In addition to having larger genomes, more transposable elements were identified in these genomes: in supergroups D and F, *wLbra*, *wWb* and *wMhie* contain more mobile elements; and *wLsig*, *wLbra* and *wMhie* more ISs (Fig. 6). Traditionally, studies of genome reduction in the cases of symbiotic bacteria indicate an expansion of mobile genetic elements in the initial stages of bacterial adaptation to a host-dependent lifestyle and an absence of mobile genetic elements in long-term obligate symbiosis associations [117]. This suggests that the different associations of *Wolbachia*-filarial nematodes represent different stages of host-dependent adaptation. Initially, it had been suggested that *Wolbachia* symbionts co-evolved with their filarial nematodes [4]. Supergroup F *Wolbachia* were thought to be the only example of horizontal transfer among the filarial nematodes [56, 73]. A recent revision of the co-phylogenetic patterns of *Wolbachia* in filariae based on multi-locus phylogenies suggests that only supergroup C *Wolbachia* exhibit strong co-speciation with their hosts [58]. Indeed, our global-fit analyses are not compatible with a global pattern of co-evolution, but rather support the hypothesis of two independent acquisitions of supergroup D and J (Fig. 4). These results highlight a differential evolution of *Wolbachia* symbiosis among the various filarial nematodes, likely having evolved from different acquisitions and subject to different selective pressures (Fig. 4, Supplementary file S3).

Another important aspect of *Wolbachia* diversity is the association between some *Wolbachia* and the WO bacteriophage [118–121]. Indeed, prophage regions have been identified in numerous *Wolbachia* genomes and the fact that these

insertions have not been eliminated by selective pressure support the hypothesis that they could provide factors of importance to *Wolbachia* [119, 122]. In the case of *Wolbachia* from arthropods, these insertions can constitute a large proportion of the *Wolbachia* genome. For example, it was recently shown that 25.4% of the *wFol* genome comprises five phage WO regions [112]. Our analyses indicate that the large-sized genomes, such as *wstri* or *wVulC*, have large regions of WO prophage (Fig. 5f, Table S9). No intact region or only vestiges of prophage regions had been observed in previously studied *Wolbachia* genomes infecting filarial nematodes [33, 43]. Our data support this absence of prophages in the newly studied genomes in this report. However, we detected some genes annotated as phage-like in the cases of *wBm*, *wBp* and *wWb* (supergroup D) and *wMhie* (supergroup F), unlike other representatives of supergroup D (*wLsig* or *wLbra*), and the genomes belonging to supergroups C and J *Wolbachia* (Table S10). Interestingly, while the bedbug symbiont *wCle* (supergroup F) has fewer phage genes than other *Wolbachia* from arthropods, numerous phage elements have been found in *Wolbachia* sequences integrated into the nuclear genome of a stronglyloidean nematode (*Dictyocaulus viviparus*) [110], which were allocated to supergroup F, suggesting significant variation in the role of phage WO within this clade. So far, *wWb*, *wBm* and *wBp* have the largest *Wolbachia* genomes of filarial nematodes, and while these phage-like insertions represent a negligible proportion of the entire genome, they nevertheless suggest that *wWb*, *wBm* and *wBp* were in contact with bacteriophages that successfully inserted DNA in their respective genomes. At the same time, our study shows that numerous genes involved in HR and the cell cycle pathway (Fig. 7) are absent in the *Wolbachia* from filarial nematodes other than *wBm*, *wBp* and *wWb*; thus, insertion of DNA might not be possible for the bacteriophages due to the nature of these genomes themselves.

Supergroup F is particularly interesting as it represents the only clade composed of *Wolbachia* symbionts of both arthropods and filarial nematodes, suggesting horizontal transfer between the two phyla [111]. Previous studies suggested it is more likely that the infection by supergroup F *Wolbachia* derived from multiple independent host switch events in the Filarioidea, because they infected species that are not closely related [56, 58]. In addition, recent phylogenomic studies suggest that supergroup F is a derived clade in the evolutionary history of *Wolbachia* [3, 20, 123]. The *wMhie* genome belonging to the supergroup F is closely related to the bedbug symbiont *wCle*; however, the characteristics of the genome (small size, few transposase elements, few phage genes, absence of prophage region) are more similar to those observed in representatives of supergroup D.

Previous genomics studies of *Wolbachia* from filarial nematodes have hypothesized mechanisms that could underpin the obligate mutualism [33, 43]. Our data indicate that both haem and nucleotide (pyrimidine and purine) metabolism are particularly conserved among all analysed *Wolbachia* genomes, even the smallest ones; thus, supporting suggestions of potential provisioning of these resources by *Wolbachia*

(Fig. 7). The hypothesis of mutualism based on nutritional provisioning has been revised after the detection of the incomplete riboflavin (vitamin B2) pathway in *wOo* [43]. Notably, the genomes of supergroup D show an incomplete folate metabolism pathway (vitamin B9) (no *folA/B/C/KP* for all representatives), which is complete for the small genomes of both supergroup C and J. The riboflavin pathway (vitamin B2) appears incomplete in supergroup C with four genes missing for *wOv*, two for *wOo* and one missing for *wDimm*, while the pathway is almost complete for the two studied supergroup J representatives (only *ribA* missing). Another interesting vitamin B pathway is the biotin operon (vitamin B7). It was previously suggested that the evolution of this operon is not congruent with proposed *Wolbachia* evolutionary history [116]. Our data show the operon is present in *wNfla* (supergroup A), *wstri*, *wLug*, *wVulC* (supergroup B), *wCle* (supergroup F) and *wCfeJ* (not belonging to a described supergroup), and that it might have been acquired horizontally as a nutritional requirement. For *wCle*, *wstri* and *wNfla*, biotin supplementation by *Wolbachia* increases insect host fitness [20, 21]. Interestingly, our study shows that incomplete metabolic pathways are not a function of *Wolbachia* genome size.

Due to their ubiquitous occurrence and diverse biological interactions with their hosts, be they in nematodes or arthropods, *Wolbachia* endosymbionts represent a striking model for studies of symbiosis. Analysis of their genomes has been used to attempt to understand the nature of the host–symbiont biochemical mechanisms, their evolutionary trajectories and their potential use for biomedical remediations. Our results pinpoint a differential evolutionary course for *Wolbachia* symbiosis among the various filarial nematode clades, suggesting evolutions from different acquisitions and subject to different selective pressures. The concept of a uniform model of symbiont–filarial host association among the *Wolbachia* clades (‘supergroups’) appears not to show consistent patterning. Overall, the pathway analysis presented in the current study suggests that no single metabolic process governs the entire spectrum of *Wolbachia*–filarial nematode associations. It is highly likely that such provisioning mechanisms might differ according to the particular host–symbiont association, although in cases where the *Wolbachia* host is itself a parasite (such as filarial nematodes), the potential metabolic interactions with mammalian and arthropod hosts of the filariae are highly complex. In the past, comparative arthropod and nematode *Wolbachia* evolutionary studies have largely been independently performed, often due to different objectives. However, understanding of the full range of diversity of *Wolbachia* genomic information will be required to comprehend their comprehensive symbiotic complexity. The analysis of the new *Wolbachia* genomes from filarial nematodes presented in the current study, as well as recent studies of *Wolbachia* genomes from arthropods more closely related to symbionts of filarial nematodes, such as the symbiont of fleas (*wCfeJ*) [92] or pseudoscorpions (*wApol*) [70], emphasize this viewpoint. Continued further genomic analyses will be instructive to highlight and help unravel these diverse symbiotic mechanisms.

Funding information

This study was supported by internal funding from New England Biolabs, except for the first Illumina library for *wDcau*, which was produced at the University of Liverpool using funds supplied by the MNHN.

Acknowledgements

We thank Andy Gardner, Clotilde Carlow, Tom Evans, Rich Roberts, Jim Ellard and Don Comb from New England Biolabs for their support. We also thank Laurie Mazzola, Danielle Fuchs and Kristen Augulewicz from the New England Biolabs Sequencing Core.

Author contributions

E. L., C. M., B. L. M., A. C. D., J. M. F. and B. E. S. conceived and designed the experiments. E. L., T. C. and B. E. S. performed the experiments. E. L. analysed the data. R. G., I. C., J. M. C.-C., K. J., N. V.-L. and C. M. provided study materials. E. L., B. E. S., J. M. F., B. L. M. and K. J. wrote the main manuscript text. All authors reviewed the manuscript.

Conflicts of interest

B. E. S. and J. M. F. are employed by New England Biolabs, Inc., who provided funding for this project.

Ethical statement

Most of the samples were collected as described in a previous study [79] and all procedures were conducted in compliance with the rules and regulations of the respective national ethical bodies. Regarding the newly studied material: the *Dirofilaria (Dirofilaria) immitis* specimen was provided by the NIAID/NIH Filariasis Research Reagent Resource Center (MTA University of Wisconsin–Oshkosh, USA; www.filariasis-center.org), and the *Litomosoides sigmodontis* specimen was provided by the MNHN, Paris, France, where the experimental procedures were carried out in strict accordance with the European Union Directive 2010/63/UE and the relevant national legislation (ethical statement no.13845).

References

- 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:e38544.
- Werren JH, Windsor DM. *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? *Proc Biol Sci* 2000;267:1277–1285.
- Brown AMV, Wasala SK, Howe DK, Peetz AB, Zasada IA et al. Genomic evidence for plant-parasitic nematodes as the earliest *Wolbachia* hosts. *Sci Rep* 2016;6:34955.
- Bandi C, Anderson TJC, Genchi C, Blaxter ML. Phylogeny of *Wolbachia* in filarial nematodes. *Proc Biol Sci* 1998;265:2407–2413.
- Werren JH, Baldo L, Clark ME. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* 2008;6:741–751.
- Kageyama D, Nishimura G, Hoshizaki S, Ishikawa Y. Feminizing *Wolbachia* in an insect, *Ostrinia furnacalis* (Lepidoptera: Crambidae). *Heredity* 2002;88:444–449.
- Bouchon D, Rigaud T, Juchault P. Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. *Proc Biol Sci* 1998;265:1081–1090.
- Duron O, Fort P, Weill M. Influence of aging on cytoplasmic incompatibility, sperm modification and *Wolbachia* density in *Culex pipiens* mosquitoes. *Heredity* 2007;98:368–374.
- Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proc Natl Acad Sci USA* 2010;107:769–774.
- Hoerauf A, Mand S, Fischer K, Kruppa T, Marfo-Debrekeye Y et al. Doxycycline as a novel strategy against bancroftian filariasis—depletion of *Wolbachia* endosymbionts from *Wuchereria bancrofti* and stop of microfilaria production. *Med Microbiol Immunol* 2003;192:211–216.
- Badawi M, Moumen B, Giraud I, Grève P, Cordaux R. Investigating the molecular genetic basis of cytoplasmic sex determination caused by *Wolbachia* endosymbionts in terrestrial isopods. *Genes* 2018;9:290.

12. Lindsey ARI, Werren JH, Richards S, Stouthamer R. Comparative genomics of a parthenogenesis-inducing *Wolbachia* symbiont. *G3* 2016;6:2113–2123.
13. Bakowski MA, McNamara CW. Advances in antiwolbachial drug discovery for treatment of parasitic filarial worm infections. *Trop Med Infect Dis* 2019;4:108.
14. Walker T, Klasson L, Sebahia M, Sanders MJ, Thomson NR et al. Ankyrin repeat domain-encoding genes in the wPip strain of *Wolbachia* from the *Culex pipiens* group. *BMC Biol* 2007;5:39.
15. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R et al. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* 2004;2:E69.
16. LePage DP, Metcalf JA, Bordenstein SR, On J, Perlmutter JI et al. Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. *Nature* 2017;543:243–247.
17. Beckmann JF, Ronau JA, Hochstrasser M. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. *Nat Microbiol* 2017;2:17007.
18. Perlmutter JI, Bordenstein SR, Unckless RL, LePage DP, Metcalf JA et al. The phage gene *wmk* is a candidate for male killing by a bacterial endosymbiont. *PLoS Pathog* 2019;15:e1007936.
19. Pichon S, Bouchon D, Liu C, Chen L, Garrett RA et al. The expression of one ankyrin *pk2* allele of the WO prophage is correlated with the *Wolbachia* feminizing effect in isopods. *BMC Microbiol* 2012;12:55.
20. Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M et al. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. *Proc Natl Acad Sci USA* 2014;111:10257–10262.
21. Ju J-F, Bing X-L, Zhao D-S, Guo Y, Xi Z et al. *Wolbachia* supplement biotin and riboflavin to enhance reproduction in planthoppers. *ISME J* 2020;14:676–687.
22. Kamtchum-Tatuene J, Makepeace BL, Benjamin L, Baylis M, Solomon T. The potential role of *Wolbachia* in controlling the transmission of emerging human arboviral infections. *Curr Opin Infect Dis* 2017;30:108–116.
23. Moreira LA, Iturbe-Ormaetxe I, Jeffery JA, Lu G, Pyke AT et al. A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, chikungunya, and *Plasmodium*. *Cell* 2009;139:1268–1278.
24. Blagrove MSC, Arias-Goeta C, Failloux A-B, Sinkins SP. *Wolbachia* strain wMel induces cytoplasmic incompatibility and blocks dengue transmission in *Aedes albopictus*. *Proc Natl Acad Sci USA* 2012;109:255–260.
25. Slatko BE, Luck AN, Dobson SL, Foster JM. *Wolbachia* endosymbionts and human disease control. *Mol Biochem Parasitol* 2014;195:88–95.
26. Bourtzis K, Dobson SL, Xi Z, Rasgon JL, Calvitti M et al. Harnessing mosquito-*Wolbachia* symbiosis for vector and disease control. *Acta Trop* 2014;132:S150–S163.
27. Ye YH, Woolfit M, Rances E, O'Neill SL, McGraw EA. *Wolbachia*-associated bacterial protection in the mosquito *Aedes aegypti*. *PLoS Negl Trop Dis* 2013;7:e2362.
28. Bouchery T, Lefoulon E, Karadjian G, Nieguitsila A, Martin C. The symbiotic role of *Wolbachia* in Onchocercidae and its impact on filariasis. *Clin Microbiol Infect* 2013;19:131–140.
29. Pfarr KM, Debrah AY, Specht S, Hoerauf A. Filariasis and lymphoedema. *Parasite Immunol* 2009;31:664–672.
30. Taylor MJ, von Geldern TW, Ford L, Hübner MP, Marsh K et al. Preclinical development of an oral anti-*Wolbachia* macrolide drug for the treatment of lymphatic filariasis and onchocerciasis. *Sci Transl Med* 2019;11:eaau2086.
31. Molyneux DH, Bradley M, Hoerauf A, Kyelem D, Taylor MJ. Mass drug treatment for lymphatic filariasis and onchocerciasis. *Trends Parasitol* 2003;19:516–522.
32. Taylor MJ, Hoerauf A, Bockarie M. Lymphatic filariasis and onchocerciasis. *Lancet* 2010;376:1175–1185.
33. Foster J, Ganatra M, Kamal I, Ware J, Makarova K et al. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol* 2005;3:e121.
34. Ghedin E, Hailemariam T, DePasse JV, Zhang X, Oksov Y et al. *Brugia malayi* gene expression in response to the targeting of the *Wolbachia* endosymbiont by tetracycline treatment. *PLoS Negl Trop Dis* 2009;3:e525.
35. Voronin D, Bachu S, Shlossman M, Unnasch TR, Ghedin E et al. Glucose and glycogen metabolism in *Brugia malayi* is associated with *Wolbachia* symbiont fitness. *PLoS One* 2016;11:e0153812.
36. Grote A, Voronin D, Ding T, Twaddle A, Unnasch TR et al. Defining *Brugia malayi* and *Wolbachia* symbiosis by stage-specific dual RNA-seq. *PLoS Negl Trop Dis* 2017;11:e0005357.
37. Li Z, Carlow CKS. Characterization of transcription factors that regulate the type IV secretion system and riboflavin biosynthesis in *Wolbachia* of *Brugia malayi*. *PLoS One* 2012;7:e51597.
38. Slatko BE, Taylor MJ, Foster JM. The *Wolbachia* endosymbiont as an anti-filarial nematode target. *Symbiosis* 2010;51:55–65.
39. Clare RH, Cook DA, Johnston KL, Ford L, Ward SA et al. Development and validation of a high-throughput anti-*Wolbachia* whole-cell screen: a route to macrofilaricidal drugs against onchocerciasis and lymphatic filariasis. *J Biomol Screen* 2015;20:64–69.
40. Xu Z, Fang S-M, Bakowski MA, Rateb ME, Yang D et al. Discovery of kirromycins with anti-*Wolbachia* activity from *Streptomyces* sp. CB00686. *ACS Chem Biol* 2019;14:1174–1182.
41. Johnston KL, Cook DAN, Berry NG, David Hong W, Clare RH et al. Identification and prioritization of novel anti-*Wolbachia* chemotypes from screening a 10,000-compound diversity library. *Sci Adv* 2017;3:eaao1551.
42. Hong WD, Benayoud F, Nixon GL, Ford L, Johnston KL et al. AWZ1066S, a highly specific anti-*Wolbachia* drug candidate for a short-course treatment of filariasis. *Proc Natl Acad Sci USA* 2019;116:1414–1419.
43. Darby AC, Armstrong SD, Bah GS, Kaur G, Hughes MA et al. Analysis of gene expression from the *Wolbachia* genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. *Genome Res* 2012;22:2467–2477.
44. Cotton JA, Bennuru S, Grote A, Harsha B, Tracey A et al. The genome of *Onchocerca volvulus*, agent of river blindness. *Nat Microbiol* 2017;2:16216.
45. Lebov JF, Mattick J, Libro S, Sparklin BC, Chung M et al. Complete genome sequence of wBp, the *Wolbachia* endosymbiont of *Brugia pahangi* FR3. *Microbiol Resour Announc* 2020;9:e00480-20.
46. Sullivan W. *Wolbachia*, bottled water, and the dark side of symbiosis. *Mol Biol Cell* 2017;28:2343–2346.
47. Comandatore F, Cordaux R, Bandi C, Blaxter M, Darby A et al. Supergroup C *Wolbachia*, mutualist symbionts of filarial nematodes, have a distinct genome structure. *Open Biol* 2015;5:150099.
48. Ramírez-Puebla ST, Servín-Garcidueñas LE, Ormeño-Orrillo E, Vera-Ponce de León A, Rosenblueth M et al. Species in *Wolbachia*? Proposal for the designation of 'Candidatus *Wolbachia bourtzisi*', 'Candidatus *Wolbachia onchocercicola*', 'Candidatus *Wolbachia blaxteri*', 'Candidatus *Wolbachia brugii*', 'Candidatus *Wolbachia taylori*', 'Candidatus *Wolbachia collem-bolicola*' and 'Candidatus *Wolbachia multihospitum*' for the different species within *Wolbachia* supergroups. *Syst Appl Microbiol* 2015;38:390–399.
49. Lindsey ARI, Bordenstein SR, Newton ILG, Rasgon JL. *Wolbachia pipientis* should not be split into multiple species: a response to Ramírez-Puebla et al., "Species in *Wolbachia*? Proposal for the designation of 'Candidatus *Wolbachia bourtzisi*', 'Candidatus *Wolbachia onchocercicola*', 'Candidatus *Wolbachia blaxteri*', 'Candidatus *Wolbachia brugii*', 'Candidatus *Wolbachia taylori*', 'Candidatus *Wolbachia collem-bolicola*' and 'Candidatus *Wolbachia multihospitum*' for the different species within *Wolbachia* supergroups". *Syst Appl Microbiol* 2016;39:220–222.

50. Chung M, Munro JB, Tettelin H, Dunning Hotopp JC. Using core genome alignments to assign bacterial species. *mSystems* 2018;3:e00236-18.
51. Newton ILG, Slatko BE. Symbiosis comes of age at the 10th biennial meeting of Wolbachia researchers. *Appl Environ Microbiol* 2019;85:e03071-18.
52. Gerth M. Classification of Wolbachia (Alphaproteobacteria, Rickettsiales): no evidence for a distinct supergroup in cave spiders. *Infect Genet Evol* 2016;43:378–380.
53. Baldo L, Werren JH. Revisiting Wolbachia supergroup typing based on WSP: spurious lineages and discordance with MLST. *Curr Microbiol* 2007;55:81–87.
54. Zhou W, Rousset F, O'Neill S. Phylogeny and PCR-based classification of Wolbachia strains using wsp gene sequences. *Proc Biol Sci* 1998;265:509–515.
55. Lo N, Casiraghi M, Salati E, Bazzocchi C, Bandi C. How many Wolbachia supergroups exist? *Mol Biol Evol* 2002;19:341–346.
56. Ferri E, Bain O, Barbuto M, Martin C, Lo N 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:e20843.
57. 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.
58. Lefoulon E, Bain O, Makepeace BL, d'Haese C, Uni S et al. Breakdown of coevolution between symbiotic bacteria *Wolbachia* and their filarial hosts. *PeerJ* 2016;4:e1840.
59. Lo N, Paraskevopoulos C, Bourtzis K, O'Neill SL, Werren JH et al. Taxonomic status of the intracellular bacterium *Wolbachia pipientis*. *Int J Syst Evol Microbiol* 2007;57:654–657.
60. Baldo L, Dunning Hotopp JC, Jolley KA, Bordenstein SR, Biber SA et al. Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Appl Environ Microbiol* 2006;72:7098–7110.
61. Casiraghi M, Bain O, Guerrero R, Martin C, Pocacqua V et al. Mapping the presence of *Wolbachia pipientis* on the phylogeny of filarial nematodes: evidence for symbiont loss during evolution. *Int J Parasitol* 2004;34:191–203.
62. Bordenstein SR, Paraskevopoulos C, Dunning Hotopp JC, Sapountzis P, Lo N et al. Parasitism and mutualism in Wolbachia: what the phylogenomic trees can and cannot say. *Mol Biol Evol* 2009;26:231–241.
63. Ma Y, Chen W-J, Li Z-H, Zhang F, Gao Y et al. Revisiting the phylogeny of *Wolbachia* in Collembola. *Ecol Evol* 2017;7:2009–2017.
64. Konecka E, Olszanowski Z. A screen of maternally inherited microbial endosymbionts in oribatid mites (Acari: Oribatida). *Microbiology* 2015;161:1561–1571.
65. Khoo JJ, Kurtti TJ, Husin NA, Beliavskaia A, Lim FS et al. Isolation and propagation of laboratory strains and a novel flea-derived field strain of *Wolbachia* in tick cell lines. *Microorganisms* 2020;8:988.
66. Werren JH, Zhang W, Guo LR. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proc Biol Sci* 1995;261:55–63.
67. Ros VID, Fleming VM, Feil EJ, Breeuwer JAJ. How diverse is the genus *Wolbachia*? Multiple-gene sequencing reveals a putatively new *Wolbachia* supergroup recovered from spider mites (Acari: Tetranychidae). *Appl Environ Microbiol* 2009;75:1036–1043.
68. Bing X-L, Xia W-Q, Gui J-D, Yan G-H, Wang X-W et al. Diversity and evolution of the *Wolbachia* endosymbionts of *Bemisia* (Hemiptera: Aleyrodidae) whiteflies. *Ecol Evol* 2014;4:2714–2737.
69. Bordenstein S, Rosengaus RB. Discovery of a novel *Wolbachia* supergroup in Isoptera. *Curr Microbiol* 2005;51:393–398.
70. Lefoulon E, Clark T, Borveto F, Perriat-Sanguinet M, Moulia C et al. Pseudoscorpion *Wolbachia* symbionts: diversity and evidence for a new supergroup S. *BMC Microbiol* 2020;20:188.
71. Sironi M, Bandi C, Sacchi L, Di Sacco B, Damiani G et al. Molecular evidence for a close relative of the arthropod endosymbiont *Wolbachia* in a filarial worm. *Mol Biochem Parasitol* 1995;74:223–227.
72. Haegeman A, Vanholme B, Jacob J, Vandekerckhove TTM, Claeys M et al. An endosymbiotic bacterium in a plant-parasitic nematode: member of a new *Wolbachia* supergroup. *Int J Parasitol* 2009;39:1045–1054.
73. Lefoulon E, Gavotte L, Junker K, Barbuto M, Uni S et al. A new type F *Wolbachia* from Splendidofiliariinae (Onchocercidae) supports the recent emergence of this supergroup. *Int J Parasitol* 2012;42:1025–1036.
74. Baldo L, Lo N, Werren JH. Mosaic nature of the *Wolbachia* surface protein. *J Bacteriol* 2005;187:5406–5418.
75. Bleidorn C, Gerth M. A critical re-evaluation of multilocus sequence typing (MLST) efforts in *Wolbachia*. *FEMS Microbiol Ecol* 2018;94:fix163.
76. Lefoulon E, Vaisman N, Frydman HM, Sun L, Volland L et al. Large enriched fragment targeted sequencing (LEFT-SEQ) applied to capture of *Wolbachia* genomes. *Sci Rep* 2019;9:5939.
77. Kent BN, Salichos L, Gibbons JG, Rokas A, Newton ILG et al. Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*) determined by targeted genome capture. *Genome Biol Evol* 2011;3:209–218.
78. Geniez S, Foster JM, Kumar S, Moumen B, LeProust E et al. Targeted genome enrichment for efficient purification of endosymbiont DNA from host DNA. *Symbiosis* 2012;58:201–207.
79. Lefoulon E, Bain O, Bourret J, Junker K, Guerrero R et al. Shaking the tree: multi-locus sequence typing usurps current Onchocercid (filarial nematode) phylogeny. *PLoS Negl Trop Dis* 2015;9:e0004233.
80. Luck AN, Anderson KG, McClung CM, VerBerkmoes NC, Foster JM et al. Tissue-specific transcriptomics and proteomics of a filarial nematode and its *Wolbachia* endosymbiont. *BMC Genomics* 2015;16:920.
81. Godel C, Kumar S, Koutsovoulos G, Ludin P, Nilsson D et al. The genome of the heartworm, *Dirofilaria immitis*, reveals drug and vaccine targets. *FASEB J* 2012;26:4650–4661.
82. Comandatore F, Sasseria D, Montagna M, Kumar S, Koutsovoulos G et al. Phylogenomics and analysis of shared genes suggest a single transition to mutualism in *Wolbachia* of nematodes. *Genome Biol Evol* 2013;5:1668–1674.
83. Saha S, Hunter WB, Reese J, Morgan JK, Marutani-Hert M et al. Survey of endosymbionts in the *Diaphorina citri* metagenome and assembly of a *Wolbachia* wDi draft genome. *PLoS One* 2012;7:e50067.
84. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH et al. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 2017;27:722–736.
85. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol* 2013;20:714–737.
86. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J et al. BLAST+: architecture and applications. *BMC Bioinformatics* 2009;10:421.
87. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 2014;30:614–620.
88. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–359.
89. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* 2018;15:461–468.
90. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
91. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUILT: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.

92. Driscoll TP, Verhoeve VI, Gillespie JJ, Johnston JS, Guillotte ML et al. A chromosome-level assembly of the cat flea genome uncovers rampant gene duplication and genome size plasticity. *BMC Biol* 2020;18:70.
93. Yoon S-H, Ha S, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek* 2017;110:1281–1286.
94. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 2013;14:60.
95. Auch AF, Klenk H-P, Göker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand Genomic Sci* 2010;2:142–148.
96. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T et al. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 2008;9:75.
97. 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:R30.
98. Arndt D, Grant JR, Marcu A, Sajed T, Pon A et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;44:W16–W21.
99. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
100. R Core Team. R: a Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2017.
101. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAA5: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 2007;35:W182–W185.
102. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 2015;16:157.
103. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 2007;56:564–577.
104. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.
105. Ioannidis P, Dunning Hotopp JC, Sapountzis P, Siozios S, Tsiamis G et al. New criteria for selecting the origin of DNA replication in Wolbachia and closely related bacteria. *BMC Genomics* 2007;8:182.
106. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M et al. Versatile and open software for comparing large genomes. *Genome Biol* 2004;5:R12.
107. Balbuena JA, Míguez-Lozano R, Blasco-Costa I. PACo: a novel procrustes application to cophylogenetic analysis. *PLoS One* 2013;8:e61048.
108. Legendre P, Desdevises Y, Bazin E. A statistical test for host-parasite coevolution. *Syst Biol* 2002;51:217–234.
109. International Helminth Genomes Consortium. Comparative genomics of the major parasitic worms. *Nat Genet* 2019;51:163–174.
110. Koutsovoulos G, Makepeace B, Tanya VN, Blaxter M. Palaeosymbiosis revealed by genomic fossils of Wolbachia in a stronglyloid nematode. *PLoS Genet* 2014;10:e1004397.
111. Casiraghi M, Bordenstein SR, Baldo L, Lo N, Beninati T et al. Phylogeny of Wolbachia pipientis based on gltA, groEL and ftsZ gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the Wolbachia tree. *Microbiology* 2005;151:4015–4022.
112. Kampfraath AA, Klasson L, Anvar SY, Vossen R, Roelofs D et al. Genome expansion of an obligate parthenogenesis-associated Wolbachia poses an exception to the symbiont reduction model. *BMC Genomics* 2019;20:106.
113. Martin C, Gavotte L. The bacteria Wolbachia in filariae, a biological Russian dolls' system: new trends in antifilarial treatments. *Parasite* 2010;17:79–89.
114. Siozios S, Ioannidis P, Klasson L, Andersson SGE, Braig HR et al. The diversity and evolution of Wolbachia ankyrin repeat domain genes. *PLoS One* 2013;8:e55390.
115. Fenn K, Blaxter M. Wolbachia genomes: revealing the biology of parasitism and mutualism. *Trends Parasitol* 2006;22:60–65.
116. Gerth M, Bleidorn C. Comparative genomics provides a time-frame for Wolbachia evolution and exposes a recent biotin synthesis operon transfer. *Nat Microbiol* 2016;2:16241.
117. McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 2011;10:13–26.
118. 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;317:1183–1188.
119. Gavotte L, Henri H, Stouthamer R, Charif D, Charlat S et al. A survey of the bacteriophage WO in the endosymbiotic bacteria Wolbachia. *Mol Biol Evol* 2007;24:427–435.
120. Masui S, Kuroiwa H, Sasaki T, Inui M, Kuroiwa T et al. Bacteriophage WO and virus-like particles in Wolbachia, an endosymbiont of arthropods. *Biochem Biophys Res Commun* 2001;283:1099–1104.
121. Wright JD, Sjöstrand 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;63:79–85.
122. Bordenstein SR, Bordenstein SR. Eukaryotic association module in phage WO genomes from Wolbachia. *Nat Commun* 2016;7:13155.
123. Gerth M, Gansauge M-T, Weigert A, Bleidorn C. Phylogenomic analyses uncover origin and spread of the Wolbachia pandemic. *Nat Commun* 2014;5:5117.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.