Multiple origins of nematode-*Wolbachia* symbiosis in supergroup F and convergent loss of bacterioferritin in filarial *Wolbachia*

A. Sinha[1], Z. Li[1], C.B. Poole[1], L. Ettwiller[1], N.F. Lima[2], M. U. Ferreira[2], F.F. Fombad[3], S. Wanji[3] and Clotilde K. S. Carlow[1*]

[1]New England Biolabs, Ipswich, Massachusetts, 01938, USA; [2]Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil; [3]Department of Microbiology and Parasitology, University of Buea, Buea, Cameroon

[*]**Author for correspondence:** Clotilde Carlow, Division of Genome Biology, New England Biolabs, Ipswich, MA 01938. Tel: 978 380 7263, Fax: 978 921 1350, e-mail: carlow@neb.com.

Data deposition

Raw Illumina reads for the *M. perstans* isolate Mpe1 and the *M. ozzardi* isolate Moz2 have been deposited in the NCBI SRA database under BioProject accession numbers PRJNA666672 and PRJNA666671 respectively. The assembled genome and PGAP annotations for the *Wolbachia* *w*Mpe and *w*Moz are available from the NCBI GenBank database with accession numbers JACZHU000000000 and JADAKK000000000 respectively. The annotated genomes for *w*Ocae and *w*MoviF are available as GenBank accessions JAHDJT000000000 and JAHRBA000000000 respectively, deposited under BioProject accessions PRJNA244005 and PRJNA739984 respectively.

Abstract

The intracellular endosymbiotic proteobacteria *Wolbachia* have evolved across the phyla nematoda and arthropoda. In *Wolbachia* phylogeny, supergroup F is the only clade with members from both arthropod and filarial nematode hosts and therefore can provide unique insights into their evolution and biology. In this study, 4 new supergroup F *Wolbachia* genomes have been assembled using a metagenomic assembly and binning approach, *w*Moz and *w*Mpe from the human filarial parasites *Mansonella ozzardi* and *Mansonella perstans*, and *w*Ocae and *w*MoviF from the blue mason bee *Osmia caerulescens* and the sheep ked *Melophagus ovinus* respectively. A comprehensive phylogenomic analysis revealed two independent origins of filarial *Wolbachia* in supergroup F, likely from ancestral arthropod hosts, as well as a convergent pseudogenization and loss of the bacterioferritin gene across all filarial *Wolbachia*. The new genomes provide a valuable resource for further studies on symbiosis, evolution, and the discovery of new antibiotics to treat mansonellosis.

Introduction

Wolbachia are gram-negative α-proteobacteria of the order *Rickettsiales*, present as intracellular symbionts in many species of parasitic filarial nematodes and arthropods[1]. While the *Wolbachia* associations in arthropods range from reproductive parasites to nutritional mutualists[1,2], in filarial nematodes, *Wolbachia* is an obligate mutualist[3] and essential for worm development, fecundity and survival[3,4]. *w*Mpe and *w*Moz are endosymbionts of *Mansonella perstans* and *Mansonella ozzardi*, respectively, the human filarial parasites responsible for mansonellosis. Mansonellosis is the most prevalent of all human filariases, yet the least studied and most neglected[5–11]. *M.*

perstans is considered to be the most common filaria in West and Central Africa[6,9], and is also found in the Amazon rainforest and the Caribbean coast of South America[9,12]. *M. ozzardi* infections have been reported from many countries in South and Central America, including some Caribbean Islands[7,9,13,14]. Co-infections of *M. perstans* and *M. ozzardi* are also known to exist in South America[15,16], complicating treatment as the two species respond differently to the commonly used anti-filarial drugs[9]. Antibiotics targeting *Wolbachia* have been successful in treating human filarial diseases such as onchocerciasis and lymphatic filariasis[4,17–22]. Doxycycline has been shown to clear *M. perstans* microfilariae[9,23,24] and may have some efficacy against *M. ozzardi*, as well as dual infections with *M. ozzardi* and *M. perstans*[16], however, no such clinical trials have been reported.

The *Wolbachia* present in the filarial nematodes of the genus *Mansonella* have a unique phylogenetic position compared to *Wolbachia* present in other human filarial parasites. The latest phylogenetic classification, based on Multi-Locus Sequence Typing (MLST), assigns *Wolbachia* to various monophyletic supergroups[25–27]. Supergroups A, B, E, H and S are comprised entirely of arthropod *Wolbachia*, while supergroups C, D and J consist of filarial *Wolbachia* exclusively. Interestingly, the *Wolbachia* from *Mansonella* are classified under supergroup F, the only supergroup to have *Wolbachia* members from nematode as well as arthropod hosts. There is considerable interest in studying these *Wolbachia* to better understand the evolution and biology of arthropod and filarial *Wolbachia*[25,26,28,29].

Genomic information provides a fundamental resource for these investigations, yet only two genomes from supergroup F are currently available, namely the genomes of *Wolbachia w*Cle from the bedbug *Cimex lectularius*[28], and *w*Mhie, from *Madathamugadia hiepei*, a filarial parasite of geckos[26]. In the current study, the genomes of supergroup F *Wolbachia*, *w*Moz and

*w*Mpe from human filarial parasites *M. ozzardi* and *M. perstans* respectively, are assembled and analyzed. The genomes of arthropod-associated supergroup F *Wolbachia*, *w*Ocae and *w*MoviF, from the mason bee, *Osmia caerulescens*, and sheep ked, *Melophagus ovinus* are also reported. Results from genome-wide analyses of synteny, phage-derived gene content, cytoplasmic incompatibility genes and metabolic pathways of these genomes are described. In addition, phylogenomic analyses of these 4 new genomes and all the available *Wolbachia* genomes were performed to generate comprehensive and robust phylogenetic trees, and confirmed the position of the new genomes within supergroup F. These analyses revealed at least two independent origins of filarial-*Wolbachia* association within supergroup F, most likely from ancestral arthropod-*Wolbachia* systems, with an accompanying loss of the bacterioferritin gene. The loss of bacterioferritin was discovered to be a feature common to all filarial *Wolbachia*.

Materials and Methods

Ethics statement

All research involving human subjects was approved by the appropriate committees and performed in accordance with all relevant guidelines and regulations. Informed consent was obtained from all participants or their legal guardians.

For *M. perstans*, ethical clearance was obtained from the National Institutional Review board, Yaoundé (REF: N° 2015/09/639/CE/CNERSH/SP) and administrative clearance from the Delegation of Public Health, South-West region of Cameroon (Re: R11/MINSANTE/SWR/ RDPH/PS/259/382). Approval for the study was granted by the “National Ethics Committee of

Research for Human Health" in Cameroon. Special consideration was taken to minimize the health risks to which any participant in this study was exposed. The objectives of the study were explained to the consenting donor after which they signed an informed consent form. The participant's documents were assigned a code to protect the privacy of the study subject. At the end of the study, the donor received a cure of mebendazole (100 mg twice daily for 30 days).

For *M. ozzardi*, study protocols were approved by the Institutional Review Board of the Institute of Biomedical Sciences, University of São Paulo, Brazil (1133/CEP, 2013). Written informed consent was obtained from all patients, or their parents or guardians if participants were minors aged <18 years. Diagnosed infections were treated with a single dose of 0.2 mg/kg of ivermectin after blood sampling[30,31].

Parasite materials and DNA extraction

M. ozzardi DNA was prepared from microfilariae collected from blood samples of individuals who participated in a previous study in Brazil[31]. A sample containing a high microfilaria load was selected for genome sequencing and is denoted as Moz1 in the present study. A Venezuelan isolate of *M. ozzardi* microfilariae was generously donated by Izaskun Petralanda in 1989 and is denoted as Moz2 in this study. Genomic DNA was prepared from Moz2 microfilariae by Proteinase K digestion followed by phenol/chloroform extraction and ethanol precipitation as well as drop dialysis (https://www.neb.com/protocols/2013/09/16/drop-dialysis) then stored at –20°C. Two independent isolates of *M. perstans* microfilariae, denoted as Mpe1 and Mpe2 in this study, were collected on nylon filters from blood samples from Cameroon[32]. Mpe1 DNA was extracted using a Genomic DNA Tissue MicroPrep kit (Zymo

Research, USA) as described previously[32]. DNA from Mpe2 was isolated as described above for Moz2. DNA quantity was determined using a Qubit dsDNA HS Assay kit in conjunction with a Qubit 2.0 Fluorometer as directed by the manufacturer (Life Technologies, USA).

Illumina library construction and sequencing

Prior to library construction, the NEBNext Microbiome DNA enrichment kit (New England Biolabs Inc., USA) was used to enrich parasite DNA and reduce human DNA contamination (except for the Moz2 sample). The preparation of Illumina libraries from Mpe1 and Moz1 samples has been described as part of a previous study[32], and a similar protocol was used for Mpe2 and Moz2, using the NEBNext Ultra II FS DNA Library Prep Kit (New England Biolabs Inc., USA) as described by the manufacturer. Following PCR amplification with different index primers to enable multiplexing, the libraries were size selected using NEBNext Sample Purification Beads (NEB cat. # E7767) following manufacturer's instructions. The approximate insert size and concentration of each library was determined using a 2100 Bioanalyzer with a high sensitivity DNA chip (Agilent Technologies, USA). Two Mpe2 libraries with insert sizes of approximately 500 and 800 bps and two Moz2 libraries with insert sizes of approximately 500 and 950 bps were constructed. Libraries were diluted to 4 nM with 10 mM Tris, 0.1 mM EDTA pH 8 for sequencing. Due to the A:T rich nature of filarial genomes, Phi X DNA was added to balance base pair composition, then sequenced using the Illumina MiSeq and NextSeq500 platforms (paired end, 150 bps).

Metagenomic assembly and binning

Raw reads were processed using the BBTools package v38.51 (https://sourceforge.net/projects/bbmap/). Duplicate raw reads and bad tiles were removed using the clumpify.sh and filterbytile.sh utilities. Adapter trimming as well as removal of phiX reads and reads with 13-nt or longer stretches of the same nucleotide (poly-G, -C, -A or-T) were performed using bbduk.sh. Human host reads were removed by mapping against the human genome (grch38) using the bbmap.sh utility. The quality metrics of the processed reads at each step were assessed using FastqC v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Reads from different runs of the libraries prepared from the same genomic DNA sample were combined and used as an input for the assembly of the metagenome using metaSpades v3.14.0[33]. Input reads were mapped back to the assembled scaffolds using bowtie2 (v.2.4.1)[34]. Binning of metagenomic scaffolds into metagenome assembled genomes of *Mansonella* and *Wolbachia* was performed using the BlobTools v1.1.1 software[35] and additional manual curation. The bins annotated as "Proteobacteria" and "Nematoda" were analyzed further to retrieve sequences genuinely originating from *Wolbachia*. A cluster of "Proteobacteria" scaffolds in the blobplots of each of the metagenomes was analyzed using blastn against the *w*Cle genome to verify their *Wolbachia* origin. These scaffolds were subsequently collected as respective *Wolbachia* assemblies from each of the isolates. The combined size of scaffolds collected from the Moz1 isolate was only 120 kb, much smaller than a 1 Mb genome typically seen for *Wolbachia*, and so were not analyzed further. Similarly, in the Mpe2 sample, only 12 scaffolds with a total size of 15 kb were identified as candidate *Wolbachia* sequences and were not analyzed further. The tight clusters of "Proteobacteria" scaffolds in the blobplots of Moz2 and

Mpe1 were classified as the *w*Moz and *w*Mpe genome assemblies respectively. For the remaining clusters of metagenome scaffolds that were also marked "Proteobacteria" by BlobTools but had a distinctly different %GC as compared to the *Wolbachia* scaffolds, blastn searches against the NCBI-nt database were performed to verify that they did not originate from the *Wolbachia* genome. All the bioinformatics programs were run using the default parameters unless otherwise stated.

The genome sequence of *Wolbachia w*Ocae from the mason bee *Osmia caerulescens* was assembled from Illumina reads available under NCBI SRA accession SRR1221705, originally described in a previous study of arthropod *Wolbachia*[36]. The reads were assembled into metagenomic scaffolds using metaSpades. A BlobTools analysis of the resulting assembly was used to identify the cluster of *Wolbachia* scaffolds.

The genome sequence of *Wolbachia w*MoviF from the sheep ked *Melophagus ovinus* was assembled from raw Illumina reads corresponding to the NCBI SRA accession number ERR969522, which originated from a study of sheep ked endosymbionts[37] that were also used recently for *Wolbachia* genome mining[38]. In the first round of assembly using all reads as input to metaSpades assembler, potential *Wolbachia* scaffolds were identified using BlobTools. However, a preliminary blastn analysis of these scaffolds indicated the presence of a supergroup A or B *Wolbachia*, in addition to the expected supergroup F *Wolbachia* in this sample. Due to the complicated nature of this dataset, additional iterations of metagenomic assembly and binning were performed using only the reads that mapped to a "bait" database containing all known *Wolbachia* genomes. The mapping of reads to bait databases was performed using bowtie2 (v.2.4.1). The bait database was updated to include the newly assembled sequences, and a new subset of reads was obtained by mapping to this updated database. These reads were then

used as input for another round of metagenomic assembly. Candidate *Wolbachia* scaffolds were selected using BlobTools. Whole genome alignments of the assembled scaffolds to three representative and complete reference genomes, namely *w*Cle (supergroup F), *w*Mel (supergroup A), and *w*AlbB (supergroup B) were performed using nucmer[39] (v4.0.0beta2) with the parameter "mincluster" set to 100.

Genome annotation and comparative analysis

For both *w*Mpe and *w*Moz genomes, as well as the *w*Ocae and *w*MoviF genomes, protein-coding genes, rRNA, tRNA, ncRNA genes and pseudogenes were identified using the NCBI prokaryotic genome annotation pipeline[40]. The completeness of the protein-coding content of all 4 genomes was assessed using the BUSCO pipeline v5.1.3 using the "proteobacteria_odb10" reference dataset[41]. Syntenic blocks between various pairs of genomes were analyzed and visualized using the JupiterPlot tool (https://github.com/JustinChu/JupiterPlot) which uses minimap2 (v 2.17-r941) for whole-genome alignments[42]. The parameters used in Jupiter plots are maxGap = 100,000; minBundleSize = 1,000; m_ref_contig_cutoff = 500; gScaff = 500. The parameter 'ng' was set to 100%, so that all the scaffolds from the two assemblies being compared are displayed in the Jupiter plots. Whole-genome alignments of *w*Moz, *w*Mpe, *w*Ocae and *w*MoviF, as well as *w*Mhie, against *w*Cle as the reference genome were performed using the nucmer utility in MUMmer4 with default parameters. The resulting alignment blocks were visualized as concentric Circos plots[43] drawn using the R package circlize[44] (v0.4.10). For global sequence comparisons across multiple *Wolbachia* genomes, the average nucleotide identity (ANI) scores

were calculated using the OrthoANIu tool[45] (v1.2). The pairwise ANI scores were used for hierarchical clustering of different *Wolbachia*, and a correlation plot was generated using R package corrplot v0.84 (https://github.com/taiyun/corrplot). For a more sensitive measure of sequence similarity and divergence between closely related *Wolbachia*, digital DNA–DNA hybridization (dDDH) scores[46] were computed using the recommended "formula 2" at the "Genome-to-Genome Distance Calculator" (GGDC) web-service (https://ggdc.dsmz.de/ggdc.php). The orthologs of protein coding genes across multiple *Wolbachia* genomes were inferred using OrthoFinder[47] (v2.4).

Insertion sequence (IS) elements were identified via the ISsaga web server[48] and by parsing the annotations in GenBank format from NCBI-PGAP pipeline. Other transposons and Group II introns were also inferred via parsing the GenBank files. Functional annotation of protein-coding genes was carried out using the eggNOG-Mapper[49] web server (http://eggnogdb.embl.de/#/app/emapper). The analysis of different metabolic pathways was based on the annotations of *w*Cle and *w*Bm genomes available in the KEGG database[50].

Phylogenomic analysis

Phylogenomic analysis was carried out on three distinct sets of input genomes. The first set comprised 127 genomes and included 123 annotated *Wolbachia* genomes available in the NCBI GenBank database (Supplementary Table 1) as well as the 4 genomes reported in this study. For a more robust and refined tree, a second set was used comprising 81 genomes selected from the 127 genomes, excluding incomplete genomes (except those within supergroup F) as well as redundant genomes e.g., multiple isolates of the same *Wolbachia* (Supplementary Table

1). The third set included only the genomes from supergroup F *Wolbachia*, and *w*CfeJ *Wolbachia* from the cat flea *Ctenocephalides felis* as an outgroup.

For each input set, OrthoFinder was used to identify Single Copy Orthologs (SCOs) conserved across all genomes within the set. Protein sequences corresponding to each orthogroup were aligned using mafft[51] v7.149b. The multiple sequence alignments were trimmed using Gblocks[52] v0.91b. The trimmed blocks from each conserved orthogroup were then concatenated into a supermatrix, and the length of each alignment block was recorded in a nexus format partition file. The corresponding gene sequence tree was generated by collecting the transcript sequences corresponding to the single copy orthologs conserved across all input *Wolbachia*. The DNA sequences corresponding to each conserved orthogroup were first aligned using mafft and then concatenated into a supermatrix sequence, with the lengths of the alignment blocks recorded in a nexus format partition file. Maximum likelihood trees were generated based on protein sequences as well as gene sequences, using the corresponding concatenated supermatrix sequence and partition files[53] as inputs to iqtree[54] v2.1.2. Automatic selection of substitution models was performed using ModelFinder[55] implementation in iqTree (iqTree command line option '-m MFP'). Evaluation of branch supports was performed using the iqTree implementations of (i) the SH-like approximate likelihood ratio test[56], and (ii) ultrafast bootstrap[57] with 1000 replicates each (iqTree command-line options '-B 1000 -alrt 1000 -abayes -lbp 1000'). The consensus trees were visualized using the iTOL webserver[58] and further annotated in Adobe Illustrator.

For a phylogenetic analysis of all supergroup F *Wolbachia* based on their 16S rRNA and ftsZ genes, corresponding sequences were obtained from GenBank. The corresponding sequences from *w*CfeJ was used as outgroup. The 16S rRNA gene sequences were aligned as

non-coding DNA in MEGAx[59] v10.2.6. The codons of ftsZ gene sequences were aligned using MUSCLE[60] in MEGAX and gap regions were removed. A maximum likelihood tree was generated from this multiple sequence alignment using iqtree online webserver W-IQ-TREE[61], with automatic selection of a codon substitution model. Bootstrap supports were calculated as described above.

Gene loss and pseudogene analysis

OrthoFinder results from the set of 81 genomes described above were analyzed to identify gene loss events across different *Wolbachia*, by cataloging orthogroups which had gene members present in all arthropod *Wolbachia* but absent in filarial *Wolbachia*.

Pseudogenization of protein-coding genes is a major factor contributing to gene loss. Pseudogenes common across different supergroup F *Wolbachia* were identified as follows. A catalog of pseudogenes for each *Wolbachia* was first obtained from its NCBI-PGAP annotation. Next, for each pseudogene, the GenBank accession for its most likely intact ancestral protein was also obtained from the NCBI-PGAP annotations. For a phylogenetically consistent and robust analysis, instead of comparing these protein accessions directly, the corresponding ancestral orthogroup was retrieved from the results of OrthoFinder and these “pseudogene-orthogroups” were compared across different *Wolbachia*. If the same orthogroup was found to be pseudogenized in more than one *Wolbachia*, the corresponding pseudogenes were designated as shared or common pseudogenes. The overlap between the ancestral pseudogene-orthogroups across different *Wolbachia* was visualized using the R package UpsetR[62]. Pseudogenes derived from transposons and IS elements were excluded from this analysis.

Annotation of phage derived elements and cytoplasmic incompatibility factors

Phage-derived genes and pseudogenes were annotated following the latest taxonomic classification scheme of *Wolbachia* phages[63]. The sequences of *Wolbachia* phage genes and pseudogenes were obtained from the Phage-WO database available at https://lab.vanderbilt.edu/bordenstein/phage-wo/database. Their best sequence matches in the proteomes of the set of 81 *Wolbachia* genomes were first identified using blastp and blastx searches. The orthogroups corresponding to these best hits were considered as phage-derived orthogroups, and the corresponding orthologs in supergroup F *Wolbachia* were annotated as phage-derived genes. Genes encoding IS-elements were excluded from this analysis to avoid mismatches and overcounting arising from the repeated and fragmented nature of genes in these families. Pseudogenes were annotated as phage-derived if their predicted ancestral proteins had an ortholog in the phage WO database. For *cifA* and *cifB* gene annotations, reference gene sequences were compiled from published data[64] and used in blast searches against the proteomes from the 81 *Wolbachia* genomes used in the phylogenomic analysis above, and the corresponding orthogroups and orthologs within supergroup F *Wolbachia* were annotated as *cifA* and *cifB*.

Results

Genomes of *w*Moz and *w*Mpe, supergroup F *Wolbachia* from human filarial parasites *M. ozzardi* and *M. perstans*

A metagenomic assembly and binning pipeline suitable for complex clinical samples was utilized to analyze sequencing data from four *Mansonella* isolates. The binning of assembled metagenome scaffolds using BlobTools identified distinct clusters of scaffolds ("blobs") corresponding to the *Wolbachia* genome, distinguishable from blobs corresponding to the genomes of the worm hosts and associated microbiota.

For the *M. ozzardi* isolate Moz1 from Brazil, only fragments of *Wolbachia* genome at low coverage were obtained and were not analyzed further (Supplementary Fig. 1a). The BlobTools analysis of *M. ozzardi* isolate Moz2 from Venezuela (Supplementary Fig. 1b) yielded a draft *w*Moz assembly of 1,073,310 bp in size comprised of 93 scaffolds. The N50 size of this assembly is 17.225 kb and the largest scaffold is 37.481 kb in length (Table 1). Of the two *M. perstans* isolates Mpe1 and Mpe2 from Cameroon, only Mpe1 yielded a *Wolbachia* assembly (Supplementary Fig. 2). The *w*Mpe assembly is comprised of 1,058,123 bp in 170 scaffolds, with a N50 size of 10.041 kb and the largest scaffold 28.485 kb in length (Table 1).

Whole genome alignment of *w*Moz and *w*Mpe demonstrated high sequence similarity and co-linearity of these independently assembled and closely related *Wolbachia* (Fig. 1). A comparison was also made to the most closely related and the only complete genome available in the Clade F supergroup, namely *w*Cle, where high sequence synteny and co-linearity were also observed (Supplementary Fig. 3).

Gene prediction using the NCBI PGAP pipeline identified 888 protein coding genes in *w*Moz and 864 protein coding genes in *w*Mpe. A BUSCO analysis (v5.1.3) was performed on these proteins to check for presence of the 219 genes conserved across most proteobacteria ("proteobacteria_odb10" database in BUSCO). The BUSCO scores of *w*Moz and *w*Mpe were 77.6% and 74.4% respectively (Table 1). These scores are typical for *Wolbachia*, even with complete genomes. For example, the BUSCO scores for *Wolbachia w*Ov from the human filarial parasite *Onchocerca volvulus*, *w*Oo from the bovine parasite *Onchocerca ochengi*, and *w*Bm from the human filarial parasite *Brugia malayi* are 76.7%, 75.8%, and 79.5% respectively (Supplementary Fig. 4). Within supergroup F, the BUSCO score for filarial *Wolbachia w*Mhie is 79%, and it is 81.7% for the arthropod *Wolbachia w*Cle.

Different isolates of *Mansonella* harbor varying levels of *Wolbachia*

The metagenomic assembly and binning approach enabled the determination of relative amounts of *Wolbachia* in different isolates by comparing read coverages of the *Mansonella*- and *Wolbachia*-specific scaffold bins in the metagenomic assemblies (Table 2). In the *M. ozzardi* sample Moz2, the median coverage of scaffolds classified as *Mansonella* was 360X, while that of the *Wolbachia* scaffolds was 13X. Assuming 1000 nematode cells per microfilaria[65], each parasite is estimated to harbor 37 *Wolbachia* cells. For the other *M. ozzardi* isolate Moz1, the median coverage for nematode scaffolds was 57X and the corresponding *Wolbachia* coverage was only 3X. Therefore, titer calculations were not performed for the Moz1 isolate as they would not be robust at such low coverage and incomplete assembly. For the *M. perstans* sample (Mpe1), the median coverage of nematode scaffolds and *Wolbachia* scaffolds were 1,032X and

30X respectively, yielding an estimated titer of 30 *Wolbachia* per microfilaria. In the other *M. perstans* isolate Mpe2, even though the median coverage for nematode scaffolds is relatively high at 633X, only 12 *Wolbachia* scaffolds were produced in the metagenomic assembly, indicating a very low titer of *Wolbachia* in this isolate. Since both isolates were collected in Cameroon, these observations point to variation in *Wolbachia* titers within the same species in a geographical area and could explain the conflicting reports of PCR-based detection of *Wolbachia* in *M. perstans*[10,23,24,66–68].

Genomes of *w*Ocae and *w*MoviF, supergroup F *Wolbachia* from the arthropods *O. caerulescens* and *M. ovinus*

For the *Wolbachia w*Ocae from the mason bee *O. caerulescens*, metagenomic assembly and binning was performed on Illumina sequencing reads available from a previous study[36]. A BlobTools analysis of the assembled scaffolds identified a distinct cluster of 186 *Wolbachia* scaffolds (Supplementary Fig. 5) with a total size of 1.2 Mb. The NCBI-PGAP pipeline identified 899 protein coding genes, which had a BUSCO score of 79% (Table 1).

The *Wolbachia* genome from the sheep ked *Melophagus ovinus* was obtained from analysis of publicly available reads[37]. A first round of metagenomic assembly and BlobTools analysis revealed a very complex mixture of *Wolbachia* as well as other proteobacterial endosymbionts (Supplementary Fig. 6). The total size of the 693 *Wolbachia* scaffolds identified in this analysis was 1.6 Mb, substantially larger than any *Wolbachia* genomes known so far. In addition, these scaffolds were spread over an unusually wide coverage range, over two orders of magnitude. A manual inspection of the blastn hits of this assembly revealed that some scaffolds

had better hits against the genome of supergroup F *Wolbachia* *w*Cle, while others were more similar to the supergroup A *Wolbachia* *w*Mel or supergroup B *Wolbachia* *w*AlbB. Together, these observations point to multiple *Wolbachia* genomes in this sample. This complexity necessitated an iterative mapping and assembly approach utilizing only the reads that mapped to *Wolbachia* genomes, which produced 572 *Wolbachia* scaffolds with a total size 1.48 Mb (Supplementary Fig. 7a). To identify the correct supergroup for each scaffold, its sequence similarity and query coverage in comparison to the *w*Cle, *w*Mel or *w*AlbB genomes were analyzed. The median percentage identity as well as query-coverage scores was observed to be the highest against the *w*Cle genome, followed by the *w*Mel genomes, and was much lower when compared to the *w*AlbB genome (Supplementary Fig. 7b, 7c). Therefore, to isolate the supergroup F scaffolds, alignments of each scaffold to *w*Cle genome was compared to its alignments against the *w*Mel genome. Among the 572 scaffolds, 40 scaffolds had a nucmer hit only to the *w*Cle genome and were therefore included in the supergroup F bin, 214 scaffolds had a hit only to the *w*Mel genome (Supplementary Fig. 7d) and were included in the supergroup A bin, while 84 scaffolds that did not have a hit in either *w*Cle of *w*Mel genome were found to be more similar to other *Wolbachia* from supergroups A or B. For the 234 scaffolds which had a nucmer hit to both *w*Cle and *w*Mel genomes, an affiliation score, calculated as the difference between its percentage identity to the *w*Cle genome and its percentage identity to the *w*Mel genome, was analyzed. This score showed a clear bimodal distribution, with a range from -15 to 13 (Supplementary Fig. 7e) confirming the presence of two *Wolbachia* from different supergroups in the host *M. ovinus*. The scaffolds which had an affiliation score greater than zero were assigned to the supergroup F. These 156 scaffolds also showed a read-coverage distribution similar to the 40 scaffolds that had exclusive hits to the *w*Cle genome (Supplementary Fig. 7f),

confirming their classification into supergroup F. The final supergroup F bin comprised of 196 scaffolds with a total size of 1.01 Mb was designated as the genome for *Wolbachia w*MoviF, where the suffix F denotes the supergroup. NCBI-PGAP annotations identified 928 protein coding genes, which had a BUSCO score of 81.3% (Table 1) and no duplicated core genes (Supplementary Fig. 4). The metagenomic bin for the putative supergroup A *Wolbachia* was highly fragmented (301 scaffolds, median length = 823 bp, total size = 0.387 Mb), and was not analyzed further.

Genome sequence comparisons between various *Wolbachia*

When all 4 new supergroup F genomes were aligned to the complete *w*Cle genome, multiple gaps were observed (Fig. 2). Interestingly, most of the regions absent from *w*Moz, *w*Mpe, *w*Ocae or *w*Mhie assemblies corresponded to regions containing IS element transposons in the *w*Cle genome (Fig. 2). This indicates that most of the protein coding regions have been recovered in these draft genome assemblies, and the missing regions mostly correspond to the IS elements, which due to their repetitive nature present a technical challenge to assemble using only short-read data[69,70].

Comparisons of sequence similarity between all supergroup F *Wolbachia* genomes, and representatives from other supergroups including those from filarial nematodes, a plant parasitic nematode, and arthropod hosts (Supplementary Table 1) were performed using genome-wide average nucleotide identity (gANI) and digital DNA–DNA hybridization (dDDH) scores. The gANI scores for both pairs of closely related sympatric *Wolbachia*, namely the *w*Oo:*w*Ov pair from *Onchocerca spp*. and the *w*Bm:*w*Bpa pair from *Brugia spp*. had gANI scores of 99% (Fig.

3a). Within supergroup F, the highest gANI scores were 97% between the *w*Moz:*w*Mpe pair and 96% between *w*MoviF:*w*Ocae pair (Fig. 3a). The filarial *Wolbachia w*Mhie was slightly more similar to the arthropod *Wolbachia w*Cle (95%) than to the filarial *Wolbachia w*Moz and *w*Mpe (92%). Comparisons between Clade F members and those of other clades generated gANI scores with a maximum value of 85% (Fig. 3a).

Sequence similarity was also analyzed using dDDH, as it provides more sensitivity when comparing closely related species [46]. The dDDH score of *w*Mpe:*w*Moz pair was found to be 71.7% (Fig. 3b). For comparison, the dDDH scores for other closely related sympatric *Wolbachia* are: 96.1 % for *w*Oo:*w*Ov pair from *Onchocerca spp*., and 93.3 % for *w*Bm:*w*Bpa pair from *Brugia spp*. (Fig. 3c, 3d). These scores suggest *w*Mpe and *w*Moz have diverged substantially from each other as compared to the *w*Oo:*w*Ov or *w*Bm:*w*Bpa pairs. Interestingly, within supergroup F, the dDDH scores (Fig. 3b) between filarial *Wolbachia w*Moz:*w*Mhie or *w*Mpe*:w*Mhie were only ~45%, lower than the score of 61.8% for the *w*Mhie:*w*Cle pair. The *w*Mpe:*w*Cle or *w*Moz:*w*Cle dDDH scores were also ~44%. Therefore, the *Wolbachia* from *Mansonella* appear to have diverged from all other supergroup F *Wolbachia*.

Orthology analysis, gene function and pathway annotations

Protein-coding gene sequences from NCBI-PGAP annotations were obtained for all 123 *Wolbachia* genomes available in GenBank (Supplementary Table 1). For orthology analysis, protein-coding genes from the 4 new genomes, in addition to 77 genomes selected to represent all supergroups, were used (see column H, Supplementary Table 1). The OrthoFinder analysis of a total 82,406 proteins encoded by these 81 genomes identified 1,751 orthogroups, 99 of which

represent Single Copy Orthologs (SCOs) conserved across all the analyzed *Wolbachia* (Supplementary Table 2). For the 4 new supergroup F *Wolbachia*, at least 96.9% of all predicted proteins could be assigned to an orthogroup (Supplementary Table 2).

Functional annotations of the proteins encoded by the *w*Moz, *w*Mpe, *w*MoviF and *w*Ocae genomes were obtained using the eggNOG database (Supplementary Table 3). Additionally, biosynthetic pathway annotations for the *w*Cle genome were obtained from the KEGG database, and corresponding orthologs were identified in the *w*Moz, *w*Mpe, *w*MoviF, *w*Ocae, and *w*Mhie proteomes. Genes encoding members of various metabolic pathways characteristic of *Wolbachia* proteomes, namely the heme pathway, purine and pyrimidine biosynthesis pathways, riboflavin metabolism, and Type IV secretion systems, were found to be mostly present in *w*Moz and *w*Mpe (Supplementary Table 4). Ribosomal protein subunits, a common target for anti-*Wolbachia* drugs such as doxycycline, as well as candidate drug targets such as pyruvate phosphate dikinase (PPDK)[71], were also present in both *w*Moz and *w*Mpe.

Biotin biosynthesis genes are absent from all supergroup F *Wolbachia* except *w*Cle

The bedbug *Wolbachia* *w*Cle genome harbors an operon encoding the enzymes involved in biosynthesis of biotin, which has been acquired via LGT from another *Rickettsia*[28]. The biotin produced by *w*Cle supports a nutritional mutualism between the bedbug and its *Wolbachia*[28]. To investigate whether this feature is common to other supergroup F *Wolbachia*, a search for the corresponding orthologs was performed. No orthologs of any of the genes involved in the biotin

biosynthesis pathway (genes *bioA*, *bioD*, *bioC*, *bioH*, *bioF* and *bioB*) could be found in *w*Moz, *w*Mpe, *w*MoviF or *w*Ocae proteomes. Additionally, when the sequencing reads from these 4 supergroup F *Wolbachia* were mapped against the *w*Cle genome, no alignments could be found within the region harboring the biotin operon, even though alignments could be obtained in the neighboring regions flanking the biotin operon (Supplementary Fig. 8). The absence of biotin pathway genes is unlikely to be due to the incomplete nature of the draft assemblies. Given the high BUSCO score (at least 79%) of the assemblies, if the biotin pathway was genuinely present in the genomes, at least fragments of a few of the 6 genes in the biotin pathway would have been detected. In contrast, the 7 genes for biosynthesis of riboflavin, another vitamin pathway important for the host-symbiont relationship[72–74] were present in *w*Cle as well as all other supergroup F *Wolbachia* (Supplementary Table 4). Thus, biotin supplementation might be unique to the bedbug and its *Wolbachia* association but is not the basis of mutualism between other supergroup F *Wolbachia* and their corresponding filarial or arthropod hosts.

Prophage remnants and mobile genetic elements in supergroup F *Wolbachia* genomes

Many *Wolbachia* are naturally infected with lambda-like temperate phages, termed WO phages, which can integrate into *Wolbachia* genomes as prophages[63]. The phage derived genes have been proposed to play important roles in generating genetic diversity, chromosomal evolution and transfer of genes required for *Wolbachia* symbiosis[63]. Phage-derived genes in all supergroup F *Wolbachia* were identified via a comprehensive search for phage-derived orthogroups (Fig. 4, Supplementary Table 5) against a recently described database specific to

Wolbachia phages[63]. In the *w*Moz and *w*Mpe genomes, 37 and 33 phage-derived orthogroups were found respectively, of which 32 were present in both *Wolbachia*. Among the gene members of these phage-derived orthogroups, "hypothetical protein" was the most common annotation (n = 22) followed by "ankyrin domain protein" (n = 12). The filarial *Wolbachia* *w*Mhie had 32 phage-associated orthogroups and shared 22 of these with *w*Moz and *w*Mpe. Neither *w*Moz, *w*Mpe, nor *w*Mhie were found to possess any core components of phage structural modules, similar to the filarial *Wolbachia* from other supergroups. Among arthropod *Wolbachia* of supergroup F, *w*Ocae had 29 phage structural genes, *w*MoviF had only 3 such genes, while *w*Cle had only one, suggesting different dynamics of phage evolution in these *Wolbachia*.

Group II introns are common, phage-derived mobile elements[63] that can be found in high copy numbers in many *Wolbachia* genomes: for example, 53 copies are found in the *w*AlbB genome[69]. Within the supergroup F genomes, the *w*Ocae genome was found to encode only two copies of Group II introns. No such genes could be detected in the *w*MoviF, *w*Moz or *w*Mpe genome assemblies. One pseudogene was detected in both *w*Moz and *w*Mpe. No intact gene or pseudogene for these elements could be found in the *w*Mhie or *w*Cle genomes, as reported earlier[26].

The *Wolbachia cifA*-*cifB* genes, which encode cytoplasmic incompatibility factors that play a key role in the reproductive biology of their arthropod hosts, are also derived from phages[63,75]. In supergroup F, no *cifA* or *cifB* genes could be detected in the genomes of arthropod *Wolbachia* *w*Cle or *w*MoviF. In *w*Ocae, multiple pseudogenes or fragments of *cifA* and *cifB* genes missing either the 5-prime or 3-prime ends were identified (Supplementary Table 6). Since all the gene fragments were observed to be located at the termini of the assembled

scaffolds, they are most likely to be pseudogenes disrupted by IS element insertions that tend to fragment the assemblies[69]. Thus, *w*Ocae also seems to be devoid of any intact and functional *cifA*-*cifB* gene pair. No *cifA* or *cifB* orthologs could be detected in filarial *Wolbachia w*Moz, *w*Mpe, or *w*Mhie.

Phylogenomic analysis encompassing all *Wolbachia* genomes

To determine the evolutionary relationships between the supergroup F *Wolbachia* and other *Wolbachia*, as well as within supergroup F, comprehensive phylogenomic analyses were performed. Using a total of 127 *Wolbachia* genomes, including 123 annotated genomes available from GenBank (Supplementary Table 1) as input for OrthoFinder, 46 Single Copy Orthologs (SCOs) that are conserved across all available genomes were identified. The corresponding phylogenetic tree based on their multiple sequence alignment spanning 8,394 amino acids was produced (Supplementary Fig. 9). For a further refinement of the tree, all incomplete or redundant genomes were excluded from analysis, except the genomes within supergroup F (see column H, Supplementary Table 1). OrthoFinder analysis on these selected 81 genomes identified 99 SCOs (Supplementary Table 2). Maximum likelihood trees with a partitioned model were built from the combined supermatrix of 62,160 nucleotides (Fig. 5) and 20,470 amino acids (Supplementary Fig. 10).. The 4 new genomes were robustly placed within supergroup F in this analysis. In this tree, the *Wolbachia w*CfeJ from the cat flea *C. felis*, which has not been assigned any supergroup yet, was found to be closest to supergroup F.

To further elucidate the phylogenetic relationships within the 6 supergroup F *Wolbachia*, additional phylogenomic analyses were performed, using *w*CfeJ as an outgroup. OrthoFinder

analyses across these 7 genomes identified 525 SCOs enabling a higher resolution of supergroup F phylogeny. The gene tree (Fig. 6) was built on a combined supermatrix of 469,845 nucleotides, and the corresponding protein tree (Supplementary Fig. 11) was derived from a combined supermatrix of 155,839 amino acids. Surprisingly, the filarial *Wolbachia* *w*Moz and *w*Mpe were placed in a clade different from the clade harboring *w*Mhie, the other filarial *Wolbachia* in this supergroup. The filarial *Wolbachia* *w*Moz and *w*Mpe were most closely related to the arthropod *Wolbachia* *w*MoviF and *w*Ocae, while filarial *Wolbachia* *w*Mhie was more closely related to the arthropod *Wolbachia* *w*Cle. Given the robust bootstrap support values for the branches (Fig. 6a), the most likely recent common ancestor for the *w*Moz, *w*Mpe lineage would be an arthropod *Wolbachia* closer to *w*MoviF and *w*Ocae, while the *w*Mhie and *w*Cle clade would have originated form a different arthropod *Wolbachia* ancestor. The clear separation of the filarial *Wolbachia* into two separate, well-supported clades reveals that the filaria-*Wolbachia* symbiosis has originated at least twice within this supergroup, from a different ancestral arthropod *Wolbachia* in each case.

To expand the phylogenetic analysis of supergroup F *Wolbachia* beyond those with whole genome sequences, all 16S rRNA and *ftsZ* gene sequences available in GenBank were utilized. The corresponding sequences from *w*CfeJ were used as an outgroup. The 16S gene tree based on 40 sequences from 15 different supergroup F *Wolbachia* had poor bootstrap support (Supplementary Fig. 12) due to the highly conserved nature of this gene. Therefore, no conclusions could be made based on this tree. The ftsZ gene tree using 32 gene sequences from 13 different *Wolbachia* had better bootstrap values (Fig. 6b), and robustly supports the separation of the *Mansonella Wolbachia* clade from the *w*Mhie clade. Interestingly, the only other filarial

Wolbachia, from the deer tick nematode *Cercopithifilaria japonica*, is placed in the same clade as *w*Mhie and *w*Cle.

Convergent loss of bacterioferritin across all filarial *Wolbachia*

Genome reduction and gene loss is a recurrent theme in the evolution of endosymbionts[76], and the genomes of filarial *Wolbachia* are often found to be even more reduced as compared to their arthropod counterparts. A study of the genes lost during this evolutionary transition can therefore reveal genes that are either potentially dispensable or are under negative selection for filaria-*Wolbachia* symbiosis. The analysis of orthology relationships across 81 *Wolbachia* genomes (Supplementary Table 2) identified various orthogroups that are absent in filarial *Wolbachia* but are present in arthropod *Wolbachia*. This approach uncovered a striking loss of the bacterioferritin gene in filarial *Wolbachia* across all supergroups C, D, J and F (Fig. 5). Within supergroup F, at least two independent losses of the bacterioferritin gene were observed, one in the *Mansonella Wolbachia* clade and the other in the *w*Mhie clade (Fig. 6a). Interestingly, pseudogene fragments of the *bfr* gene in various stages of degradation could be found across filarial *Wolbachia* (Supplementary Fig. 13a). In supergroups C and J, although *bfr* pseudogenes were not found in the NCBI-PGAP annotations, blastx searches could identify leftover fragments of bacterioferritin in the corresponding genomic region (Supplementary Fig. 13a). This genomic region was observed to be syntenic across all analyzed genomes (Supplementary Fig. 13a). Although the bacterioferritin locus in the RefSeq annotations of *Wolbachia w*Bpa from *Brugia pahangi* is annotated as an intact CDS instead of a pseudogene, a closer look at its sequence and multiple sequence alignment with other bacterioferritin genes

points to the loss of the canonical start codon ATG, similar to the corresponding sequence in *w*Bm, as well as other nucleotide insertions and deletions downstream (Supplementary Fig. 13b), suggesting the pseudogenization of this gene in *w*Bpa as well. The predicted peptide sequences corresponding to these pseudogenes showed multiple premature stop codons and frameshifts (Supplementary Fig.13c). Within supergroup F, all filarial *Wolbachia*, namely *w*Moz, *w*Mpe and *w*Mhie, contain a pseudogenized bacterioferritin, even though their most recent ancestors from arthropod *Wolbachia* have the corresponding gene intact. Since the *w*Moz:*w*Mpe pair is placed in a separate clade from *w*Mhie (Fig. 6a), our results suggest that the *bfr* gene has been pseudogenized independently in the two separate lineages.

Analysis of pseudogenes shared within the supergroup F *Wolbachia*

Intracellular bacterial symbionts, such as *Wolbachia*, are expected to accumulate pseudogenes due to genetic drift over evolutionary timescales[76]. However, the observation of two independent *bfr* pseudogenization events within supergroup F *Wolbachia* suggests that additional factors, such as convergent evolution through negative selection on certain genes could play a role in pseudogene formation. To explore the contributions of these factors, pseudogenes shared between various supergroup F *Wolbachia* were analyzed. This involved first determining the orthogroup corresponding to the ancestral gene of each pseudogene (Supplementary Table 7). These “pseudogene-orthogroups” were then compared to identify those unique or shared across different *Wolbachia* (Fig. 7). For each *Wolbachia*, the majority of its pseudogene-orthogroups were found to be unique to its genome, whereas typically very few (n = 1 to 6) were shared between different *Wolbachia*, consistent with the genetic drift model. An exception to this trend

was the number shared between the filarial *Wolbachia* *w*Moz, *w*Mpe and *w*Mhie (n = 15, Fig. 7), which included the bacterioferritin pseudogene described above. Since *w*Moz and *w*Mpe are placed in a clade separate from *w*Mhie, these 15 pseudogene-orthogroups have most likely arisen from independent, convergent loss in filarial *Wolbachia* across different lineages within supergroup F. The highest number of shared pseudogene-orthogroups was observed in the *w*Moz:*w*Mpe pair (n = 39). The corresponding pseudogenes might have arisen either through convergent evolution specific to *Wolbachia* within *Mansonella* or were formed by genetic drift in the last common ancestor of *w*Moz and *w*Mpe, or a combination of both factors.

Discussion

Wolbachia endosymbionts play critical roles in the biology of their arthropod and filarial hosts, and there is great interest in understanding their diverse functions through genomic approaches. The enigmatic supergroup F in *Wolbachia* phylogeny is the only supergroup comprising symbionts of arthropod as well as filarial nematode hosts. In this report, 4 new *Wolbachia* genomes from supergroup F are described, expanding the number of genomes from 2 to 6 in this under-represented supergroup. Two genomes are from *w*Moz and *w*Mpe, endosymbionts of *Mansonella spp.*, important filarial parasites of humans, while the other 2 genomes are from *w*Ocae and *w*MoviF, endosymbionts of the mason bee *O. caerulescens* and the sheep ked *M. ovinus* respectively.

Since *Wolbachia* are intracellular symbionts that cannot be cultured independently of their hosts, it is not possible to obtain their DNA independent of the host and the associated microbiome, presenting a challenge for accurate genome assembly. In the case of *w*Moz and

*w*Mpe, complex human clinical samples were used as the source material, containing a mixture of DNA from human blood, the filarial parasite host, its intracellular symbiont *Wolbachia*, as well as other constituents of the human microbiome. Similarly, for *w*Ocae and *w*MoviF genomes, the available sequencing data was derived from field-collected, wild isolates of their arthropod hosts, and comprised DNA from the arthropod host, *Wolbachia*, as well as the arthropod microbiome. Therefore, a metagenomic assembly and binning pipeline, customized to each of the datasets, was necessary and played a critical role in obtaining high quality *Wolbachia* genomes.

The BlobTools-based binning approach separates the assembled scaffolds based on their GC content and relative read coverage, which also provides a means for an accurate determination of *Wolbachia* titers via the measurement of relative copy numbers of genomes of *Wolbachia* versus its host. This principle has been used in a metagenomic study for *Wolbachia* titer calculations across multiple hosts[38], and for highlighting sex-specific differences in *Wolbachia* levels in *D. immitis*[77]. Given the longstanding controversy around the detection of *Wolbachia* in *M. perstans*[10,23,24,66–68], determination of *Wolbachia* titers in *Mansonella* is particularly important, and has implications for antibiotic use as a treatment[68]. In the current study, the titer in the isolate Mpe1 was found to be 30 *Wolbachia* cells per microfilaria, while a titer could not be calculated for the isolate Mpe2 due to very few and small *Wolbachia* scaffolds being recovered at low coverage, despite a high coverage obtained for the *M. perstans* genome. As both these samples were isolated in Cameroon, these observations demonstrate that even within the same geographical region, *Wolbachia* levels can vary between different isolates. For the *M. ozzardi* isolate Moz1, *Wolbachia* was detected but could not be assembled due to overall low sequencing depth obtained for this sample. In isolate Moz2, the titer was determined to be 37

Wolbachia cells per microfilaria, similar to Mpe1and levels reported from a laboratory strain of *Brugia malayi*[78]. The presence of *Wolbachia* in all *Mansonella* isolates described here, albeit at varying levels, and in another human parasite *Mansonella* spp. "DEUX"[10], together with the successful clearance of *M. perstans* microfilariae with doxycycline[24,66], supports the use of antibiotics as a promising treatment for mansonellosis.

The *Wolbachia* from supergroup F present a unique opportunity to investigate the evolutionary history of closely related arthropod and filarial *Wolbachia* since all other supergroups are comprised exclusively of either arthropod or filaria associated *Wolbachia*. A global phylogenomic analysis of 127 *Wolbachia* genomes, including all supergroup F genomes, confirmed the placement of the 4 new genomes (*w*Moz, *w*Mpe, *w*MoviF, and *w*Ocae) in supergroup F. A phylogenomic analysis focusing on all supergroup F genomes revealed two independent origins of filarial *Wolbachia* within this supergroup. The *w*Moz and *w*Mpe pair shared a common ancestor with the arthropod *Wolbachia w*Ocae and *w*MoviF, while *w*Mhie shared a different common ancestor with the bed bug *Wolbachia w*Cle. This analysis provides the first ever genomic evidence for multiple origins of filaria and *Wolbachia* symbiosis from ancestral arthropod hosts. While these evolutionary scenarios have previously been proposed based on phylogenetic studies on a handful of genes on a limited number of *Wolbachia*[79], the genome-wide phylogenies in this study are based on a supermatrix of 525 genes, incorporating 2 new filarial and 2 new arthropod *Wolbachia* genomes, thereby providing the most comprehensive and robust evidence in support of the independent origins.

Genome reduction and gene loss is a recurrent theme in the evolution of endosymbionts[76], and the genomes of filarial *Wolbachia* are often found to be even more reduced as compared to their arthropod counterparts. A study of the genes lost during this evolutionary

transition can therefore reveal genes that are dispensable for the filaria-*Wolbachia* symbiosis. In this study, gene loss was investigated using two approaches: (i) identification of protein orthogroups present in arthropod *Wolbachia* but absent from filarial *Wolbachia*; and (ii) direct comparison of annotated pseudogenes. The first approach uncovered a striking loss of the bacterioferritin gene in all filarial *Wolbachia* across all supergroups (C, D, J and F). Within supergroup F, at least two independent losses of the bacterioferritin gene were observed, one in the *w*Moz-*w*Mpe clade and the other in the *w*Mhie clade. Bacterioferritin is a critical regulator of heme homeostasis in bacteria, serving as a store for heme that can be used during heme deficiency, or sequestering excess heme to prevent toxicity[80]. In arthropods, the presence of *Wolbachia* provides a fitness benefit by increasing fecundity during conditions of iron deficiency[81], while during iron excess, *Wolbachia* protect their insect hosts from iron induced toxicity by upregulating bacterioferritin[82]. Heme is an essential nutrient required by filarial worms, but they lack a heme biosynthetic pathway. Heme provisioning by *Wolbachia* has been proposed to play a key role in the obligate host-symbiont relationship[83,84]. The lack of bacterioferritin in filarial *Wolbachia* suggests that *Wolbachia* cells can produce but no longer store heme intracellularly, making it available for uptake by the host filaria, and in doing so, affording a fitness advantage for the parasite.

In summary, the comprehensive analysis of newly assembled supergroup F *Wolbachia* genomes have provided the first ever genomic support for multiple origins of filaria and *Wolbachia* symbiosis from ancestral arthropod hosts and uncovered an intriguing loss of bacterioferritin from all filarial *Wolbachia*. Further studies of these genomes are expected to shed more light on the evolution and nature of symbiosis, as well as the identification of new treatments for mansonellosis.

Acknowledgments

The authors thank Jeremy Foster and Sofia Roitman for helpful discussions and comments on the manuscript, and Laurie Mazzola and Danielle Fuchs from the DNA sequencing core at New England Biolabs. The work was inspired by Don Comb and funded by New England Biolabs. Field research in Brazil was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), research grant to M.U.F. (2013/12723-7) and doctoral scholarship to N.F.L. (2013/ 26928-0).

Author Contributions

N.F. L., M. U. F., F.F. F., S.W. collected the clinical samples. A.S., Z. L, C.B. P., L. E. performed the sequencing experiments and bioinformatics analysis. A.S., Z. L, C.B. P., L.E., C.K.S.C. analyzed the data. A.S., Z. L, C.B. P., C.K.S.C wrote the initial manuscript. A.S., Z. L, C.B. P., L.E., M. U. F., S.W., C.K.S.C and all authors contributed to the final manuscript.

Tables

Table 1. Characteristics of genomes of *w*Moz and *w*Mpe, and other *Wolbachia* from the supergroup F

	***w*Moz**	***w*Mpe**	***w*MoviF**	***w*Ocae**	***w*Mhie**	***w*Cle**
Assembled genome size	1,073,310 bp	1,058,123 bp	1,008,858 bp	1,201,397 bp	1,025,329 bp	1,250,060 bp
Number of scaffolds	93	170	196	186	208	1
Largest scaffold	37.481 kb	28.485 kb	20.403 kb	21.794 kb	21.863 kb	-
Scaffold N50 size	17.225 kb	10.041 kb	7.443 kb	8.755 kb	6.845 kb	-
%GC	35.64	35.37	36.20	36.06	36.09	36.25
Predicted genes	1,079	1,058	1,009	1,173	1,081	1,246
Predicted proteins	888	864	928	1,055	907	981
Predicted pseudogenes	145	153	37	77	137	224
tRNA genes	36	34	35	34	30	34
BUSCO score (%)	78.1	75.3	81.3	80.4	79.0	81.7
Input	This study	This study	ERR969522 Nováková et al.	SRR122170 Gerth et al.	Lefoulon et al. 2020	Nikoh et al. 2014

reads			2015	2014		
Assembly accession and reference	GCF_020278625.1 This study	GCF_020278605.1 This study	JAHRBA000000000 This study	JAHDJT000000000 This study	GCF_013366855.1 Lefoulon et al. 2020	GCF_000829315.1 Nikoh et al. 2014

Table 2. Estimation of *Wolbachia* levels per microfilaria

Host species	**Isolate name**	**Isolate Location**	**Median coverage of Mansonella scaffolds (M)**	**Number and total size of assembled *Wolbachia* scaffolds**	**Number and total size of assembled *Wolbachia* scaffolds**	***Wolbachia* genome assembled**	**Median coverage of *Wolbachia* scaffolds (W)**	**Number o *Wolbachia* per microfilar (1000 x W/M)**
M. ozzardi	Moz1	Brazil	57	173 scaffolds	0.12 Mb	No	3	NA[#]
M. ozzardi	Moz2	Venezuela	360	93 scaffolds	1.073 Mb	*w*Moz	13	37
M. perstans	Mpe1	Cameroon	1,032	170 scaffolds	1.058 Mb	*w*Mpe	30	30
M. perstans	Mpe2	Cameroon	633	5 scaffolds	8.781 kb	No	NA[#]	NA[#]

***** Assuming 1000 cells per microfilaria

[#] Not calculated due to the very low number or coverage of *Wolbachia* scaffolds

Fig. Legends

Fig. 1. High sequence similarity and synteny between the *de novo* assembled *w*Moz and *w*Mpe genomes

The *Wolbachia* genomes assembled from two biologically and geographically distinct isolates display a high level of synteny and inter-genome consistency. The genomes were aligned using minimap2 and visualized using the JupiterPlot software. The *w*Moz scaffolds are displayed as colored boxes in the left semi-circle, while the *w*Mpe scaffolds are shown as gray boxes in the right semi-circle. The syntenic regions are marked by bands connecting scaffolds from one assembly to the corresponding region in the other assembly. Regions of rearrangement are shown as bands crisscrossing the horizontal bands of the syntenic regions.

Fig. 2. Assembled genomes include most of the protein coding regions of *Wolbachia*, with the gaps corresponding to IS elements.

Whole genome alignments of *w*Mpe (blue ring), *w*Moz (light-blue ring), *w*Ocae (green ring), *w*MoviF (light green ring) and *w*Mhie (innermost gray ring) to the *w*Cle chromosome (outermost black circle) are visualized as a circos plot. The white bars mark the locations of the IS elements in the *w*Cle genome. The gaps in the alignments of various genomes to the *w*Cle genome overlap extensively with IS elements in the *w*Cle genome.

Fig. 3. Global comparisons between multiple *Wolbachia* genomes using genome-wide Average Nucleotide Identity (gANI) scores, and digital DNA:DNA Hybridization (dDDH) scores.

(a) gANI scores were calculated between pairs of *Wolbachia* from filarial and plant-parasitic nematodes, as well as the arthropod *Wolbachia w*Cle. Hierarchical clustering of pairwise gANI scores, presented as a correlation matrix, ordered the various *Wolbachia* in a pattern recapitulating their supergroup assignments. The red lines mark the boundaries between clusters representing different supergroups. (b) dDDH scores between members of supergroup F (c) dDDH scores between members of supergroup C, and (d) dDDH scores between members of supergroup D.

Fig. 4. Distribution of phage-derived gene functions in supergroup F *Wolbachia*

Phage derived genes in each *Wolbachia* were categorized according to their predicted protein function. Only *w*Ocae has intact phage modules, while *w*MoviF and *w*Cle have only one and two core structural phage genes respectively. Phage structural genes were absent from filarial *Wolbachia w*Moz, *w*Mpe and *w*Mhie.

Fig. 5. Global phylogenomic analysis of all *Wolbachia* genomes

Phylogenomic tree based on gene sequences of 91 single copy orthologs (SCOs) conserved across 81 *Wolbachia* representing all supergroups. Supergroups comprised of *Wolbachia* exclusively from arthropod hosts are marked in blue, while the supergroups comprised exclusively of filarial *Wolbachia* are marked in orange. The supergroup F, which has members

from both arthropod and filarial hosts, is marked in purple background. The red asterisks denote the *Wolbachia* with a pseudogenized or absent bacterioferritin gene.

Fig. 6. Two independent origins of *Wolbachia*-filaria symbiosis within supergroup F

(a) The phylogenomic tree based on gene sequences of 525 SCOs conserved across all supergroup F *Wolbachia* indicates at least two independent origins of filaria-*Wolbachia* symbiosis, most likely from an arthropod-*Wolbachia* ancestor. The *w*Mpe/*w*Moz lineage has split from the clade containing *w*Ocae and *w*MoviF, while *w*Mhie shares a different common ancestor with the bed bug *Wolbachia w*Cle. The arthropod *Wolbachia w*CfeJ from the cat flea *C. felis* was used as the outgroup. The red asterisks denote the *Wolbachia* with a pseudogenized or absent bacterioferritin gene. The bootstrap values are shown along the respective branches. (b) Phylogenetic tree of supergroup F *Wolbachia* based on ftsZ gene sequence also supports at least two independent origins of filarial *Wolbachia*. All *Wolbachia* from *Mansonella* hosts, including *M. perforata* are placed in a clade adjacent to the clade harboring arthropod *Wolbachia w*Ocae and *w*MoviF, while the *Wolbachia w*Mhie and *w*Cjaponica from filarial hosts *M. hiepei* and *Cercopithifilaria japonica* are in a separate clade, more closely related to the arthropod *Wolbachia w*Cle. The bootstrap values are shown along the respective branches. Filarial *Wolbachia* are marked with orange font.

Fig. 7. Pseudogenes shared between supergroup F *Wolbachia*

The orthogroups corresponding to the ancestral gene of each pseudogene were compared across different *Wolbachia*. Each *Wolbachia* harbors its unique set of pseudogenes, reflecting their origin as evolutionary accidents. The largest set of shared pseudogenes is observed between the closely related *w*Moz and *w*Mpe *Wolbachia* (n = 39, dark orange bar), suggesting either a shared

common ancestor or convergent evolution within their respective *Mansonella* hosts. The next largest shared set of pseudogenes is found among the *Wolbachia w*Moz, *w*Mpe, and *w*Mhie (n = 15, blue bar), which are phylogenetically in different clades, indicating that the corresponding genes have undergone convergent loss of function across the three filarial *Wolbachia* in supergroup F. Transposons such as IS elements were excluded from this analysis due to their tendency to frequently convert into pseudogenes.

Supplementary Information

Supplementary Fig. 1. BlobTools based binning of the metagenomes assembled from *Mansonella ozzardi* isolates (a) Moz1, from Brazil (b) Moz2, from Venezuela

Each dot in the plot represents a scaffold in the assembled metagenome, plotted by its %GC content on the x-axis and log10 of its coverage on the y-axis. The dot size is scaled according to log10 of the length of each scaffold. Colors were assigned based on taxonomic annotations from BlobTools and manual annotations. Bins that were finally assigned to the *Mansonella* host and the *Wolbachia* endosymbiont are marked by enclosing boxes. Boxes with dashed borders indicate locations of expected *Wolbachia* scaffolds, but very few or very small scaffolds were recovered for these samples.

Supplementary Fig. 2. BlobTools based binning of the metagenomes assembled from *Mansonella perstans* isolates (a) Mpe1, and (b) Mpe2, both from Cameroon

Each dot in the plot represents a scaffold in the assembled metagenome, plotted by its %GC content on the x-axis and log10 of its coverage on the y-axis. The dot size is scaled according to log10 of the length of each scaffold. Colors were assigned based on taxonomic annotations from

BlobTools and manual annotations. Bins that were finally assigned to the *Mansonella* host and the *Wolbachia* endosymbiont are marked by enclosing boxes. Boxes with dashed borders indicate locations of expected *Wolbachia* scaffolds, but very few or very small scaffolds were recovered for these samples.

Supplementary Fig. 3. Jupiter plots show regions of conservation and synteny between (a) *w*Moz and (b) *w*Mpe in comparison to the *w*Cle genome

The *w*Cle chromosome is represented in linear form as the blue semi-circle on the left in both panels. The gray boxes represent scaffolds of (a) *w*Moz draft genome, and (b) *w*Mpe draft genome. The horizontal bands connect the conserved regions across the two *Wolbachia*. Regions of re-arrangements are shown as bands crisscrossing the horizontal bands of syntenic regions.

Supplementary Fig. 4. BUSCO scores of newly assemble supergroup F *Wolbachia* as compared to other related *Wolbachia*

The BUSCO pipeline was used to measure the proportion of highly conserved, single copy orthologs (BUSCO groups) across multiple *Wolbachia* genomes, including filarial Wolbachia and all supergroup F *Wolbachia*. The set of reference BUSCO groups was set to the lineage "proteobacteria_odb10", which contains 219 BUSCOs derived from proteobacterial species. The complete and circular *w*Oo genome has the lowest BUSCO score, marked by a vertical dotted line. The supergroups are indicated within a parenthesis after the corresponding *Wolbachia* names. The BUSCO scores of the *Wolbachia* genomes reported in this study are comparable to other *Wolbachia* genomes, indicating high quality of these draft genomes.

Supplementary Fig. 5. Taxonomic binning of metagenomic scaffolds to identify *Wolbachia* from the mason bee *Osmia caerulescens*

The cluster of scaffolds identified as *Wolbachia* by the BlobTools analysis are tightly clustered and well separated from the scaffolds from the arthropod host.

Supplementary Fig. 6. Multiple endosymbionts in the complex microbiome of *Melophagus ovinus*

(a) A BlobPlot showing the scaffolds from the first round of metagenomic assembly. The scaffolds corresponding to a few selected, high abundance taxa, namely the arthropod host, *Wolbachia* and other proteobacterial endosymbionts *Arsenophonus*, *Bartonella* and *Sodalis*, and a Euglenozoan member of the host microbiome. (b) Same Blobplot as in (b) but showing only the scaffolds corresponding to various endosymbionts, highlighting the complexity of this sample, which necessitated a bait-and-assemble approach.

Supplementary Fig. 7. *Melophagus ovinus* has a supergroup F *Wolbachia* as well as a supergroup A *Wolbachia*

(a) BlobPlot showing the metagenome assembled scaffolds of *Wolbachia* based on a bait-and-assemble approach. Briefly, a database of all the known genomes of *Wolbachia* were used as bait to fish out reads mapping to these genomes. These filtered reads were then used as input for metagenomic assembly. The BlobPlot indicates a wide range of read coverage for *Wolbachia* scaffolds, atypical for samples which have only one *Wolbachia*. (b) Boxplots showing distribution of nucmer-based percent identity of each scaffold when aligned to *w*Cle, *w*Mel or *w*AlbB genome. (c) Boxplots showing distribution of percentage of length of each query scaffold when aligned to *w*Cle, *w*Mel or *w*AlbB genomes using nucmer. (d) A Venn diagram

showing the distribution of the assembled *Wolbachia* scaffolds with hits to supergroup A genome *w*Mel, supergroup F *w*Cle or other *Wolbachia*. (e) The *Wolbachia* scaffolds from the metagenome assembly were assigned an affiliation score measuring the difference in alignment score of each scaffold against a supergroup F genome *w*Cle versus a supergroup A *Wolbachia* *w*Mel. Affiliation score for each scaffold was measured as the difference in its percentage identity to the *w*Cle genome versus the *w*Mel genome. The histogram as well as a density plot of these affiliation scores shows a clear bimodal distribution, demonstrating the presence of two different *Wolbachia*, and the utility of the affiliation score in separating the scaffolds originating from two different supergroup *Wolbachia*. (f) Read-coverage distributions of scaffolds in different *Wolbachia* bins. Each dot represents a scaffold. The bins are labeled on the x-axis, and the read-coverage is represented on the y-axis.

Supplementary Fig. 8. No evidence for biotin biosynthesis genes in any supergroup F *Wolbachia* except *w*Cle

Raw sequencing reads from various supergroup F *Wolbachia* do not show any alignments within biotin operon of *w*Cle, even though they have alignments mapping to the flanking regions.

Supplementary Fig. 9. Phylogenomic tree of 127 *Wolbachia* genomes based on 46 Single Copy Orthologs

All 123 *Wolbachia* genomes and their annotated proteomes that were available in the NCBI Genbank or RefSeq databases were downloaded (in May 2021) for this analysis. The proteomes were used as input to the OrthoFinder software, which also generated the phylogenomic tree using default parameters. The concatenated supermatrix used for the multiple sequence alignment was 8,394 amino acids in length.

Supplementary Fig. 10. Phylogenomic tree of 81 *Wolbachia* genomes based on 99 Single Copy Orthologs

Tree based on the combined supermatrix of 99 proteins conserved across 81 *Wolbachia* genomes was analyzed. Total length of the concatenated supermatrix is 20,470 amino acids. Substitution model selection was performed by ModelFinder, and applied to each gene separately, using a partition model. Branch support was based on 1,000 bootstrap replicates and values greater than 70 are displayed in the tree. The supergroups with exclusively arthropod hosts are marked with a blue background, and supergroups with exclusively filarial hosts are marked with an orange background. Supergroup F is marked with a purple background.

Supplementary Fig. 11. Phylogenomic tree of supergroup F *Wolbachia* genomes based on 525 Single Copy Orthologs

Tree based on the combined supermatrix of 525 proteins conserved across 6 supergroup F *Wolbachia* and the outgroup *w*CfeJ. Total length of the concatenated supermatrix is 155,39 amino acids. Substitution model selection was performed by ModelFinder, applied to each gene separately, using a partition model. Branch support was based 1,000 bootstrap replicates and values greater than 70 are displayed on the tree. Red asterisks indicate *Wolbachia* from filarial parasites.

Supplementary Fig. 12. Phylogenomic tree of supergroup F *Wolbachia* genomes based on the 16s rRNA gene

The 16s rRNA gene sequences were aligned using MUSCLE in MEGA-X, and the tree was inferred by iqtree online server. The alignment consisted of 40 sequences from 15 different *Wolbachia* with 1774 columns and 368 distinct patterns. Only 71 sites are parsimony-

informative (610 singleton sites, 1093 constant sites), leading to very short branch lengths with poor bootstrap support. Filarial *Wolbachia* are highlighted in red font. Bootstrap values are in blue font.

Supplementary Fig. 13. Convergent loss of bacterioferritin gene across all available filarial *Wolbachia* genomes

(a) Multiple sequence alignment of the genomic region flanking the bacterioferritin gene *bfr* shows conserved syntenic genes. The sequences are grouped according to their supergroup. Arthropod *Wolbachia* sequences are labelled as such. Annotated *bfr* pseudogenes are highlighted in magenta. Regions of tblastn hit to a particular *Wolbachia* genome are indicated with purple arrows. No tblastn hit could be found in the genome of the filarial *Wolbachia w*Dcau, even though the rest of the genomic region was syntenic with other *Wolbachia*. (b) Multiple sequence alignment of the DNA sequence in the *bfr* gene region for all filarial *Wolbachia*, with arthropod *Wolbachia w*Cle, *w*Ocae, *w*MoviF and *w*CfeJ sequences as references for an intact *bfr* gene. (c) Multiple sequence alignment of the predicted protein / ORF of the *bfr* region across all filarial *Wolbachia*, with the arthropod bacterioferritin protein sequences included as reference for a complete ORF. Premature stop codons in each amino acid sequence are shown as asterisks.

Supplementary Table 1. *Wolbachia* genomes used for comparative and phylogenomic analyses

Supplementary Table 2: Orthology analysis using OrthoFinder

(a) Orthogroups and their member genes for different *Wolbachia* (b) Genes not assigned to any orthogroups. (c) Overall OrthoFinder statistics for each *Wolbachia* genome analyzed.

Supplementary Table 3: eggNOG annotations of proteins in (a) *w*Moz and (b) *w*Mpe (c) *w*Ocae (d) *w*MoviF genomes

Supplementary Table 4: Various biological pathways in *w*Moz, *w*Mpe and other *Wolbachia* in supergroup F

(a) Heme pathway (b) Purine (c) Pyrimidine (d) Type IV and Type VI secretion systems

Supplementary Table 5: Annotations of phage-derived genes in supergroup F *Wolbachia*

(a) Annotations of phage-derived genes in all supergroup F *Wolbachia*. The database of phage-derived gene sequences available at https://lab.vanderbilt.edu/bordenstein/phage-wo/database/ was used to query the OrthoFinder results from 81 *Wolbachia* genomes described in Supplementary Table 2. The items in the column "Phage_element_type" and "Phage_element_subType" follow the annotation scheme described in Bordenstein and Bordenstein 2021. (b) Distribution of phage-derived genes into different functional categories for each supergroup F *Wolbachia*

Supplementary Table 6: Annotations of *cifA-cifB* genes in *w*Ocae

Supplementary Table 7: Summary of pseudogenes common between supergroup F *Wolbachia*

Pseudogenes and their corresponding ancestral proteins were identified from NCBI-PGAP annotations for (a) *w*Moz (b) *w*Mpe (c) *w*MoviF (d) *w*Ocae (e) *w*Cle, and (f) *w*Mhie genomes. The orthogroups for the ancestral gene, "pseudogene-orthogroups", were identified from the OrthoFinder results. The pseudogenes which derived from IS elements, marked by column "Transposon" set as "TRUE", were excluded from the analysis of shared pseudogene-orthogroups. (g) List of non-transposon pseudogene-orthogroups shared across supergroup F *Wolbachia*.

References

1. Werren, J. H., Baldo, L. & Clark, M. E. *Wolbachia*: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* **6**, 741–751 (2008).
2. Zug, R. & Hammerstein, P. Bad guys turned nice? A critical assessment of *Wolbachia* mutualisms in arthropod hosts. *Biol. Rev.* **90**, 89–111 (2015).
3. Taylor, M. J., Bandi, C. & Hoerauf, A. *Wolbachia* Bacterial Endosymbionts of Filarial Nematodes. in *Advances in Parasitology* (eds. Baker, J. R., Muller, R. & Rollinson, D.) vol. 60 245–284 (Academic Press, 2005).
4. Pfarr, K. M. & Hoerauf, A. M. Antibiotics which target the *Wolbachia* endosymbionts of filarial parasites: a new strategy for control of filariasis and amelioration of pathology. *Mini Rev. Med. Chem.* **6**, 203–210 (2006).
5. Downes, B. & Jacobsen, K. A systematic review of the epidemiology of mansonelliasis. *Afr. J. Infect. Dis.* **4**, (2010).
6. Simonsen, P. E., Onapa, A. W. & Asio, S. M. *Mansonella perstans* filariasis in Africa. *Acta Trop.* **120**, S109–S120 (2011).
7. Lima, N. F., Aybar, C. A. V., Juri, M. J. D. & Ferreira, M. U. *Mansonella ozzardi*: a neglected New World filarial nematode. *Pathog. Glob. Health* **110**, 97–107 (2016).
8. Mediannikov, O. & Ranque, S. Mansonellosis, the most neglected human filariasis. *New Microbes New Infect.* **26**, S19–S22 (2018).
9. Ta-Tang, T.-H., Crainey, J. L., Post, R. J., Luz, S. L. & Rubio, J. M. Mansonellosis: current perspectives. *Res. Rep. Trop. Med.* **9**, 9–24 (2018).
10. Sandri, T. L. *et al.* Molecular Epidemiology of *Mansonella* Species in Gabon. *J. Infect. Dis.* **223**, 287–296 (2021).

11. Bélard, S. & Gehringer, C. High Prevalence of *Mansonella* Species and Parasitic Coinfections in Gabon Calls for an End to the Neglect of *Mansonella* Research. *J. Infect. Dis.* **223**, 187–188 (2021).
12. Tavares da Silva, L. B. *et al*. Molecular Verification of New World *Mansonella perstans* Parasitemias. *Emerg. Infect. Dis.* **23**, 545–547 (2017).
13. Raccurt, C. P. *Mansonella ozzardi* and its vectors in the New World: an update with emphasis on the current situation in Haiti. *J. Helminthol*. **92**, 655–661 (2018).
14. Calvopina, M., Chiluisa-Guacho, C., Toapanta, A., Fonseca, D. & Villacres, I. High Prevalence of *Mansonella ozzardi* Infection in the Amazon Region, Ecuador. **25**, (2019).
15. Kozek, W. J., Palma, G., Henao, A., García, H. & Hoyos, M. Filariasis in Colombia: Prevalence and Distribution of *Mansonella Ozzardi* and *Mansonella (=Dipetalonema) Perstans* Infections in the Comisaría Del Guainía. *Am. J. Trop. Med. Hyg*. **32**, 379–384 (1983).
16. Crainey, J. L. *et al.* Deep-sequencing reveals occult mansonellosis co-infections in residents from the Brazilian Amazon village of São Gabriel da Cachoeira. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* (2020) doi:10.1093/cid/ciaa082.
17. Langworthy, N. G. *et al.* Macrofilaricidal activity of tetracycline against the filarial nematode *Onchocerca ochengi*: elimination of *Wolbachia* precedes worm death and suggests a dependent relationship. *Proc. R. Soc. Lond. B Biol. Sci*. **267**, 1063–1069 (2000).
18. Bazzocchi, C. *et al*. Combined ivermectin and doxycycline treatment has microfilaricidal and adulticidal activity against *Dirofilaria immitis* in experimentally infected dogs. *Int. J. Parasitol.* **38**, 1401–1410 (2008).

19. Clare, R. H. *et al.* Industrial scale high-throughput screening delivers multiple fast acting macrofilaricides. *Nat. Commun.* **10**, 11 (2019).
20. Taylor, M. J. *et al.* Preclinical development of an oral anti-*Wolbachia* macrolide drug for the treatment of lymphatic filariasis and onchocerciasis. *Sci. Transl. Med.* **11**, (2019).
21. Hübner, M. P. *et al.* In vivo kinetics of *Wolbachia* depletion by ABBV-4083 in *L. sigmodontis* adult worms and microfilariae. *PLoS Negl. Trop. Dis.* **13**, e0007636 (2019).
22. Hong, W. D. *et al.* AWZ1066S, a highly specific anti-*Wolbachia* drug candidate for a short-course treatment of filariasis. *Proc. Natl. Acad. Sci.* **116**, 1414–1419 (2019).
23. Keiser, P. B. *et al.* Molecular identification of *Wolbachia* from the filarial nematode *Mansonella perstans*. *Mol. Biochem. Parasitol.* **160**, 123–128 (2008).
24. Coulibaly, Y. I. *et al.* A Randomized Trial of Doxycycline for *Mansonella perstans* Infection. *N. Engl. J. Med.* **361**, 1448–1458 (2009).
25. Lefoulon, E. *et al.* Breakdown of coevolution between symbiotic bacteria *Wolbachia* and their filarial hosts. *PeerJ* **4**, e1840 (2016).
26. Lefoulon, E. *et al.* Diminutive, degraded but dissimilar: *Wolbachia* genomes from filarial nematodes do not conform to a single paradigm. *Microb. Genomics* **6**, e000487 (2020).
27. Lefoulon, E. *et al.* Pseudoscorpion *Wolbachia* symbionts: diversity and evidence for a new supergroup S. *BMC Microbiol.* **20**, 188 (2020).
28. Nikoh, N. *et al.* Evolutionary origin of insect–*Wolbachia* nutritional mutualism. *Proc. Natl. Acad. Sci.* **111**, 10257–10262 (2014).
29. Hosokawa, T., Koga, R., Kikuchi, Y., Meng, X.-Y. & Fukatsu, T. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proc. Natl. Acad. Sci.* **107**, 769–774 (2010).

30. Basano, S. de A. *et al.* Phase III Clinical Trial to Evaluate Ivermectin in the Reduction of *Mansonella ozzardi* infection in the Brazilian Amazon. *Am. J. Trop. Med. Hyg.* **98**, 786–790 (2018).
31. Lima, N. F., Gonçalves-Lopes, R. M., Kruize, Y. C. M., Yazdanbakhsh, M. & Ferreira, M. U. CD39 and immune regulation in a chronic helminth infection: The puzzling case of *Mansonella ozzardi*. *PLoS Negl. Trop. Dis.* **12**, e0006327 (2018).
32. Poole, C. B. *et al.* In Silico Identification of Novel Biomarkers and Development of New Rapid Diagnostic Tests for the Filarial Parasites *Mansonella perstans* and *Mansonella ozzardi*. *Sci. Rep.* **9**, 10275 (2019).
33. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
34. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
35. Laetsch, D. R. & Blaxter, M. L. BlobTools: Interrogation of genome assemblies. *F1000Research* **6**, 1287 (2017).
36. Gerth, M., Gansauge, M.-T., Weigert, A. & Bleidorn, C. Phylogenomic analyses uncover origin and spread of the Wolbachia pandemic. *Nat. Commun.* **5**, 5117 (2014).
37. Nováková, E., Husník, F., Šochová, E. & Hypša, V. Arsenophonus and Sodalis Symbionts in Louse Flies: an Analogy to the Wigglesworthia and Sodalis System in Tsetse Flies. *Appl. Environ. Microbiol.* **81**, 6189–6199 (2015).
38. Scholz, M. *et al.* Large scale genome reconstructions illuminate Wolbachia evolution. *Nat. Commun.* **11**, 5235 (2020).

39. Marçais, G. *et al*. MUMmer4: A fast and versatile genome alignment system. *PLOS Comput. Biol.* **14**, e1005944 (2018).

40. Tatusova, T. *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624 (2016).

41. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

42. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

43. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).

44. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. *Bioinforma. Oxf. Engl.* **30**, 2811–2812 (2014).

45. 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* **110**, 1281–1286 (2017).

46. Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P. & Göker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**, 60 (2013).

47. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

48. Varani, A. M., 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.* **12**, R30 (2011).

49. Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).

50. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).

51. Nakamura, T., Yamada, K. D., Tomii, K. & Katoh, K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **34**, 2490–2492 (2018).

52. Talavera, G. & Castresana, J. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Syst. Biol.* **56**, 564–577 (2007).

53. Chernomor, O., von Haeseler, A. & Minh, B. Q. Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. *Syst. Biol.* **65**, 997–1008 (2016).

54. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

55. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

56. Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

57. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

58. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

59. Stecher, G., Tamura, K. & Kumar, S. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Mol. Biol. Evol.* **37**, 1237–1239 (2020).

60. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

61. Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–W235 (2016).

62. Conway, J. R., Lex, A., Gehlenborg, N. & Hancock, J. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).

63. Bordenstein, S. R. & Bordenstein, S. R. The most widespread phage in animals: Genomics and taxonomic classification of Phage WO. *bioRxiv* 2021.11.04.467296 (2021) doi:10.1101/2021.11.04.467296.

64. Martinez, J., Klasson, L., Welch, J. J. & Jiggins, F. M. Life and Death of Selfish Genes: Comparative Genomics Reveals the Dynamic Evolution of Cytoplasmic Incompatibility. *Mol. Biol. Evol.* **38**, 2–15 (2021).

65. Basyoni, M. M. A. & Rizk, E. M. A. Nematodes ultrastructure: complex systems and processes. *J. Parasit. Dis.* **40**, 1130–1140 (2016).

66. Debrah, L. B. *et al*. The Efficacy of Doxycycline Treatment on *Mansonella perstans* Infection: An Open-Label, Randomized Trial in Ghana. *Am. J. Trop. Med. Hyg.* **101**, 84–92 (2019).

67. Grobusch, M. P., Kombila, M., Autenrieth, I., Mehlhorn, H. & Kremsner, P. G. No evidence of *Wolbachia* endosymbiosis with *Loa loa* and *Mansonella perstans*. *Parasitol. Res.* **90**, 405–408 (2003).

68. Gehringer, C. *et al*. Molecular Evidence of *Wolbachia* Endosymbiosis in *Mansonella perstans* in Gabon, Central Africa. *J. Infect. Dis.* **210**, 1633–1638 (2014).

69. Sinha, A., Li, Z., Sun, L. & Carlow, C. K. S. Complete Genome Sequence of the *Wolbachia w*AlbB Endosymbiont of Aedes albopictus. *Genome Biol. Evol.* **11**, 706–720 (2019).

70. Lefoulon, E. *et al*. Large Enriched Fragment Targeted Sequencing (LEFT-SEQ) Applied to Capture of *Wolbachia* Genomes. *Sci. Rep.* **9**, 1–10 (2019).

71. Raverdy, S., Foster, J. M., Roopenian, E. & Carlow, C. K. S. The *Wolbachia* endosymbiont of *Brugia malayi* has an active pyruvate phosphate dikinase. *Mol. Biochem. Parasitol.* **160**, 163–166 (2008).

72. Li, Z. & Carlow, C. K. S. Characterization of Transcription Factors That Regulate the Type IV Secretion System and Riboflavin Biosynthesis in *Wolbachia* of *Brugia malayi*. *PLOS ONE* **7**, e51597 (2012).

73. Moriyama, M., Nikoh, N., Hosokawa, T. & Fukatsu, T. Riboflavin Provisioning Underlies Wolbachia's Fitness Contribution to Its Insect Host. *mBio* **6**, (2015).

74. Ju, J.-F. *et al.* Wolbachia supplement biotin and riboflavin to enhance reproduction in planthoppers. *ISME J.* **14**, 676–687 (2020).

75. Lindsey, A. R. I. *et al.* Evolutionary Genetics of Cytoplasmic Incompatibility Genes cifA and cifB in Prophage WO of *Wolbachia*. *Genome Biol. Evol.* **10**, 434–451 (2018).

76. Andersson, S. G. E. & Kurland, C. G. Reductive evolution of resident genomes. *Trends Microbiol.* **6**, 263–268 (1998).

77. Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M. & Blaxter, M. Blobology: exploring raw genome data for contaminants, symbionts, and parasites using taxon-annotated GC-coverage plots. *Bioinforma. Comput. Biol.* **4**, 237 (2013).

78. McGarry, H. F., Egerton, G. L. & Taylor, M. J. Population dynamics of *Wolbachia* bacterial endosymbionts in *Brugia malayi*. *Mol. Biochem. Parasitol.* **135**, 57–67 (2004).

79. Lefoulon, E. *et al.* A new type F *Wolbachia* from Splendidofilariinae (Onchocercidae) supports the recent emergence of this supergroup. *Int. J. Parasitol.* **42**, 1025–1036 (2012).

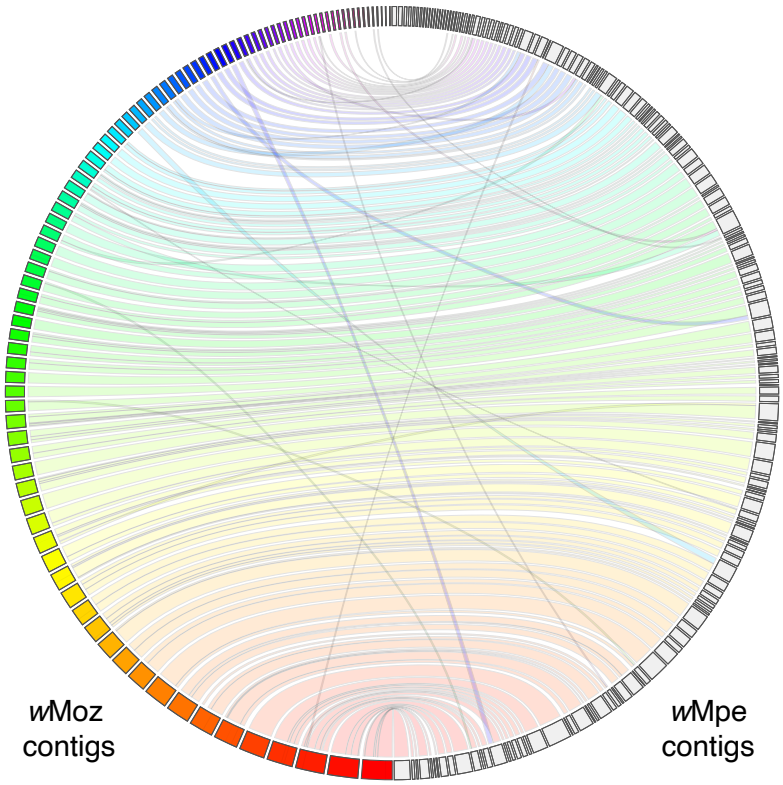
80. Carrondo, M. A. Ferritins, iron uptake and storage from the bacterioferritin viewpoint. *EMBO J.* **22**, 1959–1968 (2003).

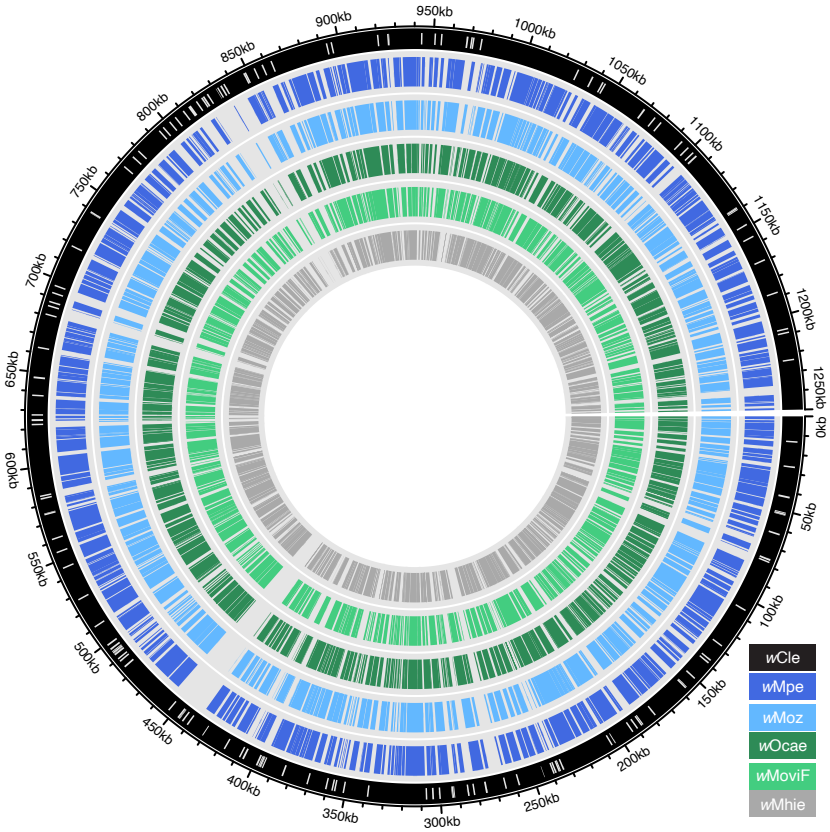
81. Brownlie, J. C. *et al.* Evidence for Metabolic Provisioning by a Common Invertebrate Endosymbiont, Wolbachia pipientis, during Periods of Nutritional Stress. *PLOS Pathog.* **5**, e1000368 (2009).

82. Kremer, N. *et al.* Wolbachia Interferes with Ferritin Expression and Iron Metabolism in Insects. *PLOS Pathog.* **5**, e1000630 (2009).

83. Foster, J. *et al.* The *Wolbachia* Genome of *Brugia malayi*: Endosymbiont Evolution within a Human Pathogenic Nematode. *PLOS Biol.* **3**, e121 (2005).

84. Gill, A. C., Darby, A. C. & Makepeace, B. L. Iron Necessity: The Secret of Wolbachia's Success? *PLoS Negl. Trop. Dis.* **8**, e3224 (2014).


*w*Moz
contigs
*w*Mpe
contigs


0kb
50kb
100kb
150kb
200kb
250kb
300kb
350kb
400kb
450kb
500kb
550kb
600kb
650kb
700kb
750kb
800kb
850kb
900kb
950kb
1000kb
1050kb
1100kb
1150kb
1200kb
1250kb
wCle
wMpe
wMoz
wOcae
wMoviF
wMhie

a

		wPpe	wMoz	wMpe	wMoviF	wOcae	wCle	wMhie	wLsig	wWb	wBm	wBpa	wCtub	wDcau	wDimm	wOo	wOv
L	wPpe	100															
F	wMoz	73	100														
F	wMpe	73	97	100													
F	wMoviF	74	92	93	100												
F	wOcae	73	92	92	96	100											
F	wCle	73	92	92	94	93	100										
F	wMhie	73	92	92	94	93	95	100									
D	wLsig	74	83	83	83	83	84	83	100								
D	wWb	74	84	84	84	84	85	85	89	100							
D	wBm	74	83	83	84	84	84	84	89	97	100						
D	wBpa	74	83	83	84	84	84	84	89	97	99	100					
J	wCtub	73	80	80	80	81	81	80	80	80	80	80	100				
J	wDcau	74	80	80	81	81	81	80	80	80	80	80	82	100			
C	wDimm	73	81	81	82	82	82	82	81	81	81	81	83	83	100		
C	wOo	73	82	81	82	82	82	82	81	81	81	82	83	83	86	100	
C	wOv	73	81	82	82	82	82	82	81	81	81	81	83	83	86	99	100

b

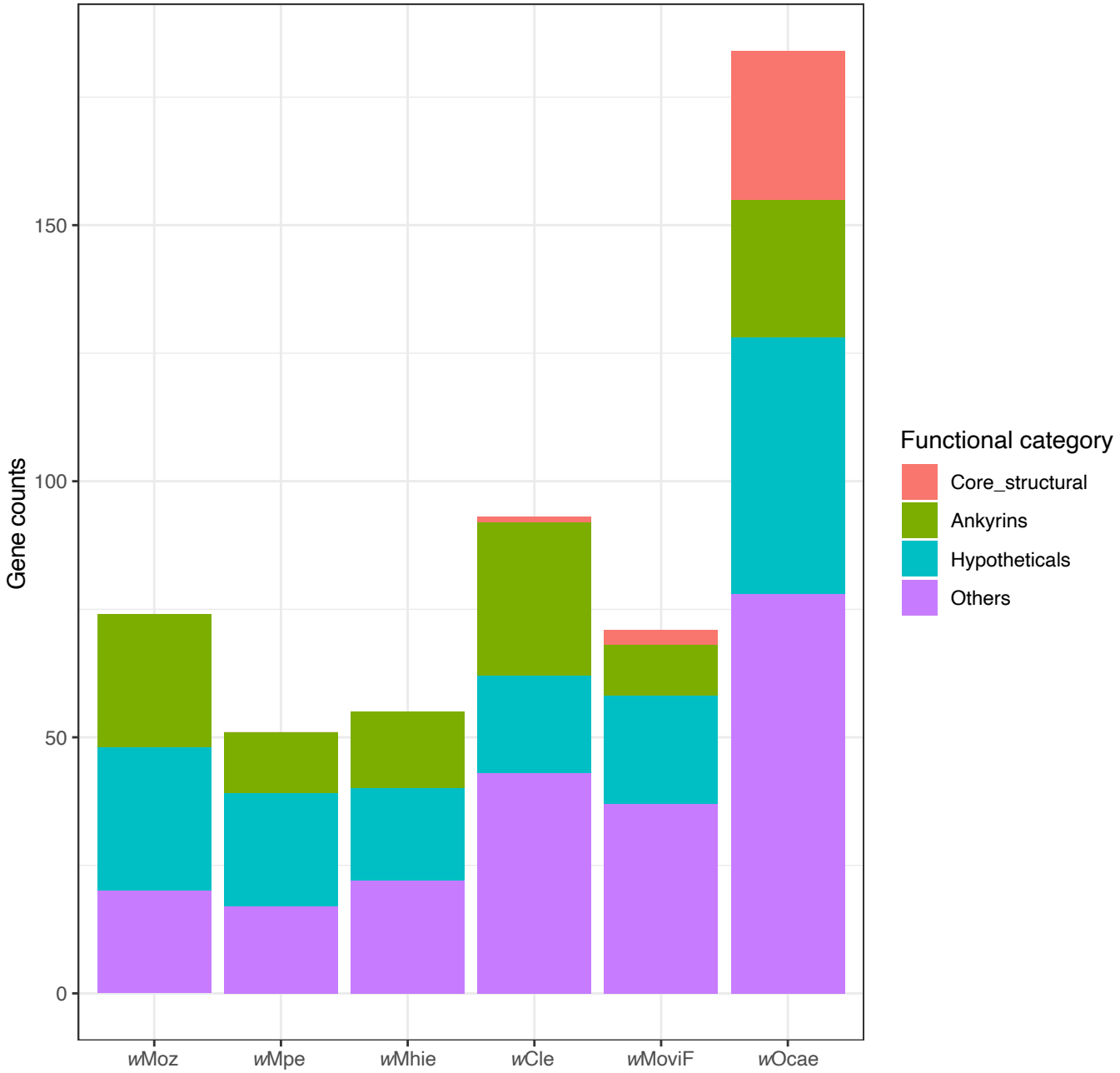
	wMoz	wMpe	wMoviF	wOcae	wCle
wMoz					
wMpe	**71.7** [68.7 - 74.5]				
wMoviF	47.4 [44.8 - 50]	47.9 [45.3 - 50.5]			
wOcae	45.4 [42.9 - 48]	46.3 [43.7 - 48.8]	66.6 [63.6 - 69.4]		
wCle	43.6 [41.1 - 46.2]	44.1 [41.5 - 46.6]	53.9 [51.2 - 56.6]	50.0 [47.4 - 52.6]	
wMhie	45.3 [42.7 - 47.9]	45.7 [43.1 - 48.2]	54.1 [51.4 - 56.8]	52.3 [49.7 - 55]	61.8 [58.9 - 64.6]

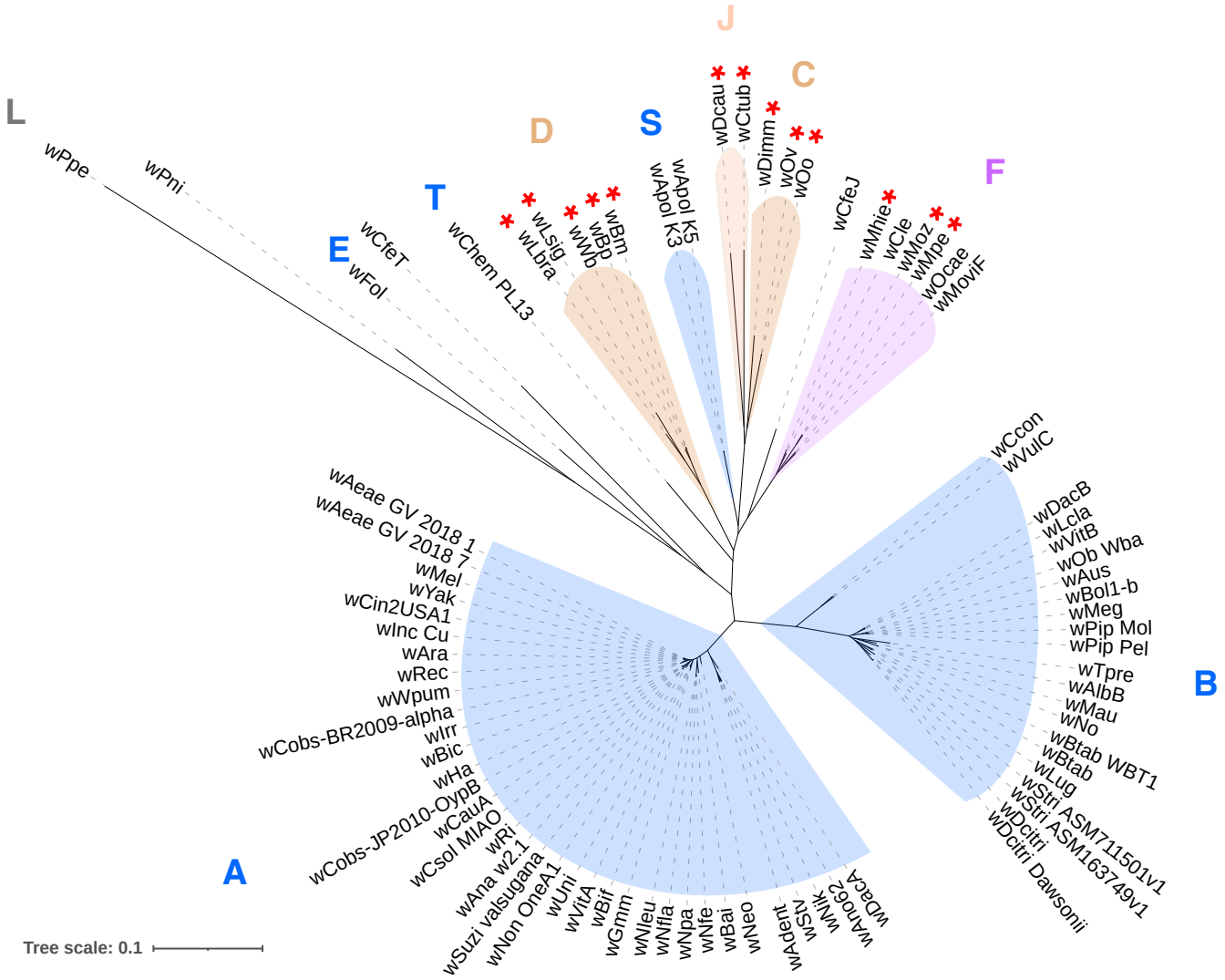
c

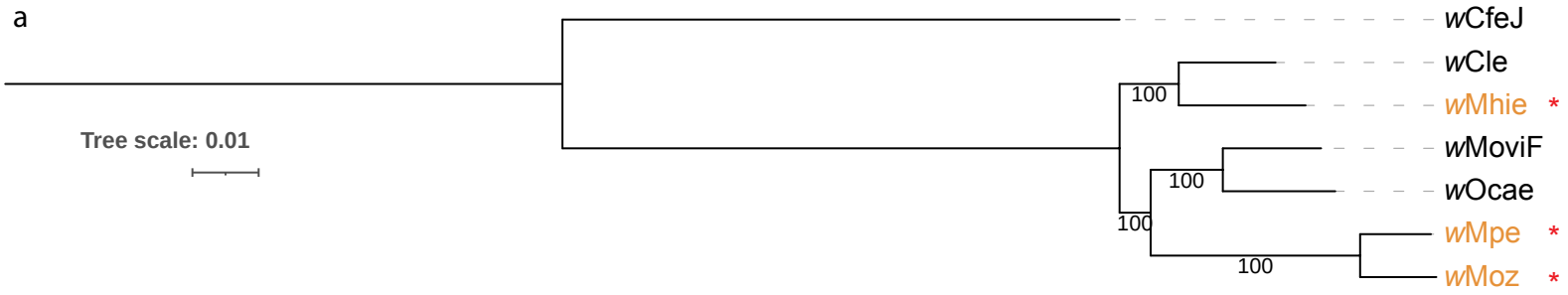
	wOo	wOv
wOo		
wOv	**96.1** [94.8 - 97.2]	
wDimm	29.2 [26.8 - 31.7]	29.1 [26.7 - 31.6]

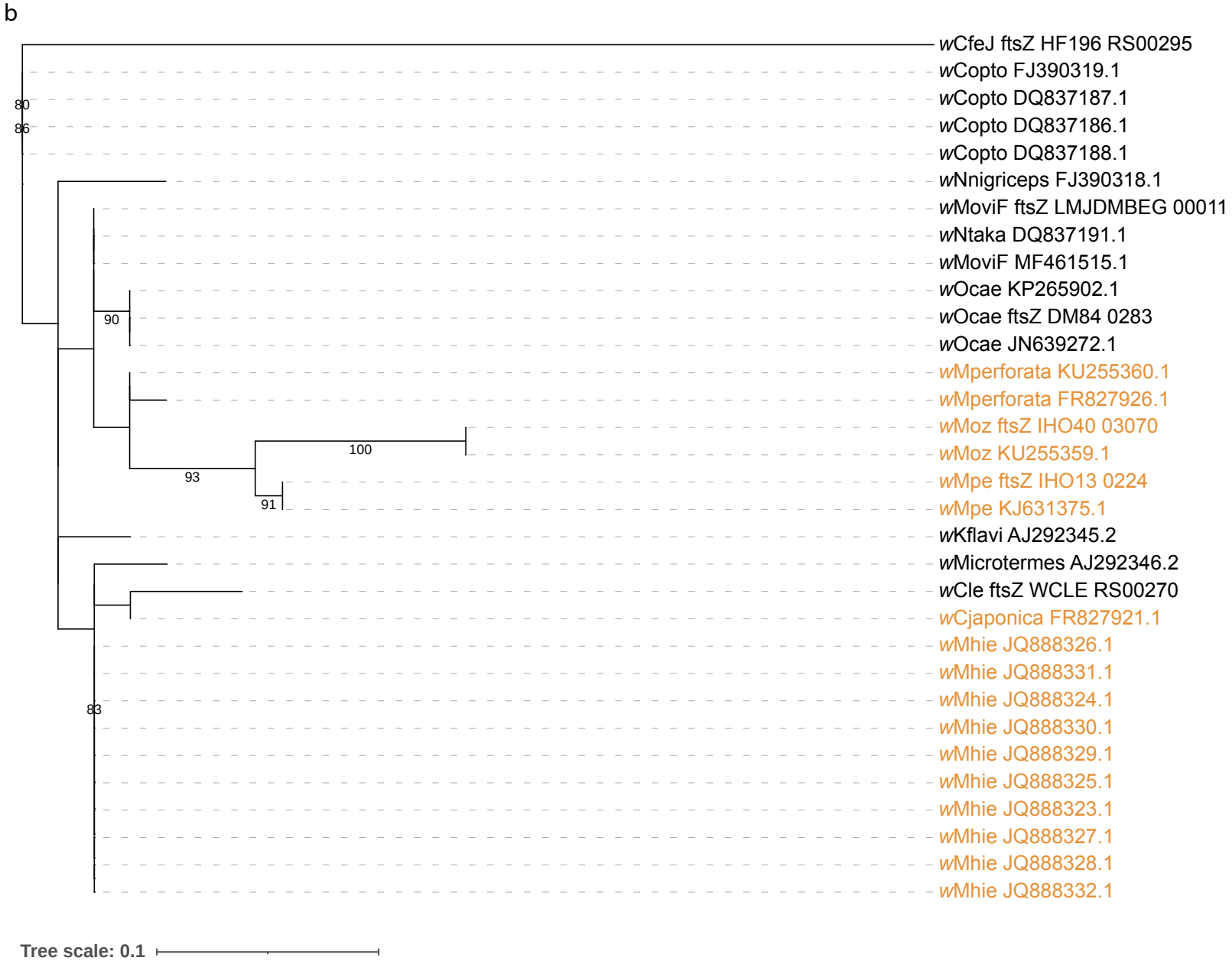
d

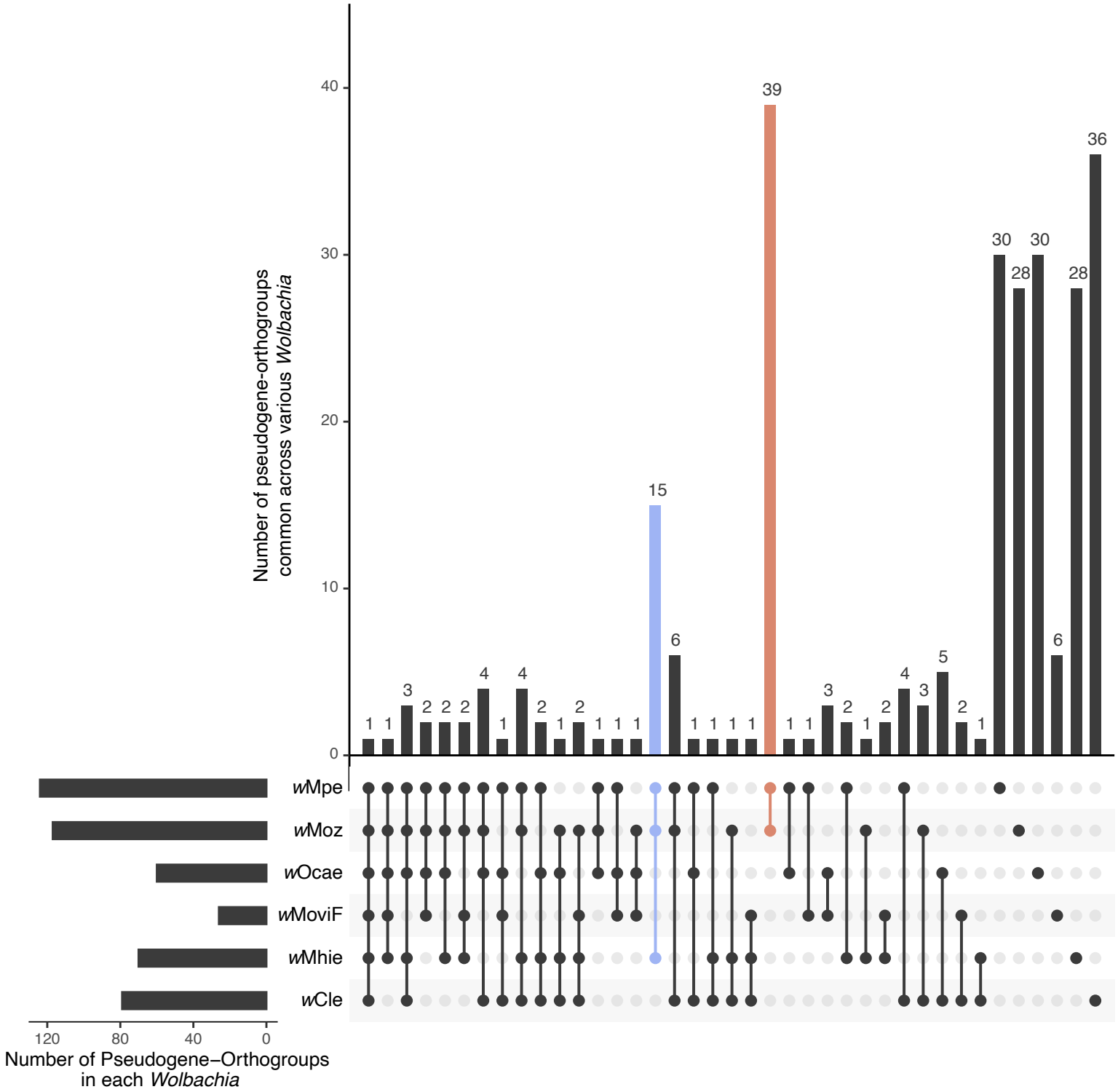
	wBm	wBp	wWb
wBm			
wBp	**93.3** [91.4 - 94.8]		
wWb	70.7 [67.7 - 73.5]	70.3 [67.3 - 73.1]	
wLsig	36.5 [34 - 39]	36.4 [33.9 - 38.9]	36.8 [34.3 - 39.3]


Gene counts
150
100
50
0
wMoz
wMpe
wMhie
wCle
wMoviF
wOcae
Functional category
Core_structural
Ankyrins
Hypotheticals
Others


L
wPpe
wPni
E
wFol
wCfeT
T
wChem PL13
D
wLbra
wLsig
wWb
wBp
wBm
S
wApol K3
wApol K5
J
wDcau
wCtub
C
wDimm
wOv
wOo
wCfeJ
F
wMhie
wCle
wMoz
wMpe
wOcae
wMoviF
wCcon
wVulC
B
wDacB
wLcla
wVitB
wOb Wba
wAus
wBol1-b
wMeg
wPip Mol
wPip Pel
wTpre
wAlbB
wMau
wNo
wBtab WBT1
wBtab
wLug
wStri ASM711501v1
wStri ASM163749v1
wDcitri
wDcitri Dawsonii
wAeae GV 2018 1
wAeae GV 2018 7
wMel
wYak
wCin2USA1
wInc Cu
wAra
wRec
wWpum
wCobs-BR2009-alpha
wIrr
wBic
wHa
wCobs-JP2010-OypB
wCauA
wCsol MIAO
wRi
wAna w2.1
wSuzi valsugana
wNon OneA1
wUni
wVitA
wBif
wGmm
wNleu
wNfla
wNpa
wNfe
wBai
wNeo
wApent
wSiv
wMik
wAno62
wDaca
A
Tree scale: 0.1


a
wCfeJ
wCle
wMhie *
wMoviF
wOcae
wMpe *
wMoz *
100
100
100
100
Tree scale: 0.01


b
wCfeJ ftsZ HF196 RS00295
wCopto FJ390319.1
wCopto DQ837187.1
wCopto DQ837186.1
wCopto DQ837188.1
wNnigriceps FJ390318.1
wMoviF ftsZ LMJDMBEG 00011
wNtaka DQ837191.1
wMoviF MF461515.1
wOcae KP265902.1
wOcae ftsZ DM84 0283
wOcae JN639272.1
wMperforata KU255360.1
wMperforata FR827926.1
wMoz ftsZ IHO40 03070
wMoz KU255359.1
wMpe ftsZ IHO13 0224
wMpe KJ631375.1
wKflavi AJ292345.2
wMicrotermes AJ292346.2
wCle ftsZ WCLE RS00270
wCjaponica FR827921.1
wMhie JQ888326.1
wMhie JQ888331.1
wMhie JQ888324.1
wMhie JQ888330.1
wMhie JQ888329.1
wMhie JQ888325.1
wMhie JQ888323.1
wMhie JQ888327.1
wMhie JQ888328.1
wMhie JQ888332.1
80
86
90
100
93
91
83
Tree scale: 0.1


Number of pseudogene-orthogroups common across various Wolbachia
0
10
20
30
40
1 1 3 2 2 2 4 1 4 2 1 2 1 1 1 15 6 1 1 1 1 39 1 1 3 2 1 2 4 3 5 2 1 30 28 30 6 28 36
wMpe
wMoz
wOcae
wMoviF
wMhie
wCle
120
80
40
0
Number of Pseudogene−Orthogroups in each Wolbachia