Genomes of *Wolbachia* endosymbionts from the human filarial parasites *Mansonella perstans* and *Mansonella ozzardi*

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

Running Title: Genomes of *Wolbachia* from *Mansonella spp.*

[*]**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.

Abstract

Mansonella ozzardi and *Mansonella perstans*, filarial parasites infecting millions of people worldwide, harbor their unique obligate endosymbionts, the alpha-proteobacteria *Wolbachia* *w*Moz and *w*Mpe, respectively. Currently, little is known about these *Wolbachia* and no genome sequences are available. In the current study, high quality draft genomes of *w*Moz and *w*Mpe were assembled from complex clinical samples using a metagenome assembly and binning approach. These represent the first genomes from supergroup F *Wolbachia* originating from human parasites and share features characteristic of filarial as well arthropod *Wolbachia*, consistent with their position in supergroup F. Metagenomic data analysis was also used to estimate *Wolbachia* titers, which revealed wide variation in levels across different clinical isolates, addressing the contradicting reports on presence or absence of *Wolbachia* in *M. perstans*. These findings may have implications for the use antibiotics to treat mansonellosis. The *w*Moz and *w*Mpe genome sequences provide a valuable resource for further studies on symbiosis, evolution and drug discovery.

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], *Wolbachia* is an obligate mutualist in the filarial nematodes[3] and is essential for worm development, fecundity and survival[3,4]. The latest phylogenetic classification, based on Multi-Locus Sequence Typing (MLST), assigns *Wolbachia* to various monophyletic supergroups[5–7]. 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* present in the filarial nematodes of the genus *Mansonella* are classified under supergroup F, the only supergroup to have *Wolbachia* members from nematode as well as arthropod hosts.

Mansonella perstans and *Mansonella ozzardi* are the two main causes of mansonellosis, and harbor their specific *Wolbachia*, *w*Mpe and *w*Moz respectively. Mansonellosis is the most prevalent human filariases, yet the least studied and most neglected of all filarial infections[8–13]. *M. perstans* is considered to be the most common filaria in West and Central Africa[9,12]. It is also found in the Amazon rainforest and the Caribbean coast in South America[12,14]. *M. ozzardi* infections have been reported from many countries in South and Central America, including some Caribbean Islands[10,12,15,16]. Importantly, co-infections of *M. perstans* and *M. ozzardi* have been reported in South America[17,18], making their treatment challenging as the two species respond differently to the commonly used anti-filarial drugs[12]. Antibiotics targeting *Wolbachia* have been successful in treating filarial diseases such as onchocerciasis and lymphatic filariasis[4,19–24]. Doxycycline, an effective anti-*Wolbachia* antibiotic, has been shown to clear *M. perstans* microfilariae[12,25–27]. Although no clinical trials have been reported, it is likley that

antibiotics will have a similar effect against *M. ozzardi*[28], enabling the use of a single drug to treat co-infections with *M. ozzardi* and *M. perstans*[18].

There is great interest in understanding the nature of the symbiotic relationship between *Mansonella* and their *Wolbachia* endosymbionts. The phylogenetic position of the *Mansonella Wolbachia* in a unique supergroup that also contains arthropod *Wolbachia* could also have implications for the nature of interactions with their host[5,6,29,30]. While genomic information could provide fundamental resource for investigating their biology, the only complete genome from supergroup F *Wolbachia* is the genome of *Wolbachia* from the bedbug *Cimex lectularius* (*w*Cle)[29]. A draft genome of the supergroup F *Wolbachia* from *Madathamugadia hiepei* (*w*Mhie), a filarial parasite of geckos, has become available recently[6]. However, since no genomic information is available for the *Wolbachia* of *Mansonella*, it is not known whether their biology is more closely aligned to filarial or arthropod *Wolbachia*[12].

Here, the first genomes of supergroup F *Wolbachia w*Moz and *w*Mpe from human filarial parasites are reported. These draft genomes have been assembled from one isolate of *M. ozzardi* microfilariae from Venezuela, and one isolate of *M. perstans* from Cameroon. The genomes are slightly larger than 1 Mb and encode for about 900 genes each. The genome drafts are of high quality based on assessment of completeness using presence of conserved *Wolbachia* core genes as a metric. Analysis of deep-sequencing data from additional isolates of *M. ozzardi* and *M. perstans* was performed and provided some insights into the longstanding controversy around the status of *Wolbachia* across different *Mansonella perstans* isolates[25–27,31–33]. Additionally, comparative analyses between *w*Moz and *w*Mpe, as well as across other filarial and arthropod *Wolbachia* genomes are also performed and reported.

The availability of the *w*Moz and *w*Mpe genome sequences will provide further insight into the evolution and phylogenetic relationships of the *Wolbachia* endosymbionts of old and new world *Mansonella* species. They will serve as an important resource for further studies on symbiosis, facilitate the development of improved methods for *Wolbachia* detection and contribute to the on-going search for new anti-*Wolbachia* drugs.

Materials and Methods

Ethics statement

All research involving human subjects was approved by the appropriate committee 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 given 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 a day 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[34,35].

Parasite materials and DNA extraction

DNA from *M. ozzardi* microfilariae from blood samples of individual subjects were collected as part of a previous study in Brazil[35]. One of the high microfilaremia samples 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, and drop dialysis (https://www.neb.com/protocols/2013/09/16/drop-dialysis) then stored at –20ºC. Two independent isolates of *M. perstans* microfilariae, denoted as Mpe1 and Mpe2 in this study, were obtained on nylon filters from blood samples of consenting individuals in Cameroon[36]. Mpe1 DNA was extracted using a Genomic DNA Tissue MicroPrep kit (Zymo Research, USA) as described previously[37]. 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

The NEBNext Microbiome DNA enrichment kit (New England Biolabs Inc., USA) was used as directed to enrich for parasite DNA and reduce human DNA contamination prior to construction of the libraries from each of the samples, except Moz2. The preparation of Illumina libraries from Mpe1 and Moz1 samples have been described as part of a previous study[37]. A similar protocol was used for preparing libraries from Mpe2 and Moz2. Briefly, Illumina libraries were constructed 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://jgi.doe.gov/data-and-tools/bbtools/). Duplicate raw reads and bad tiles were removed using the clumpify.sh and filterbytile.sh utilities. Adapter trimming, removal of phiX reads and reads with 13-nt or longer stretches of the same nucleotide (poly-G, -C, -A or-T) was 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[38]. Input reads were mapped back to the assembled contigs using bowtie2 (v.2.4.1)[39]. Binning of metagenomic contigs into metagenome assembled genomes (MAGs) of *Mansonella* and *Wolbachia* was performed using the BlobTools v1.1.1 software[40] and additional curation. The bins annotated as "Proteobacteria" and "Nematoda" were analyzed further to retrieve sequences originating from *Wolbachia*. A cluster of "Proteobacteria" contigs in the blobplots of each of the metagenomes was analyzed further using blastn against *w*Cle genome to verify their *Wolbachia* origin. These contigs were subsequently collected as respective *Wolbachia* assemblies from each of the isolates. The combined size of contigs collected from the Moz1 isolate was only 120 kb, much smaller than a 1 Mb genome typically seen for *Wolbachia*, and were not analyzed further. Similarly, in the Mpe2 sample, only 12 contigs with a total size of 15 kb were identified as candidate *Wolbachia* sequence and were not analyzed further. The tight clusters of "Proteobacteria" contigs in the blobplots of Moz2 and Mpe1 were analyzed further as the *w*Moz and *w*Mpe genome assemblies respectively. For the remaining clusters of metagenome contigs that were marked also "Proteobacteria" by BlobTools but had a distinctly different %GC as compared to the *Wolbachia* contigs, blastn searches against the NCBI-nt database was 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.

Genome annotation and comparative analysis

For both *w*Mpe and *w*Moz genomes, protein-coding genes, rRNA, tRNA, ncRNA genes and pseudogenes were identified using the NCBI prokaryotic genome annotation pipeline[41]. Analysis of sequence similarity and synteny between *w*Moz, *w*Mpe and *w*Cle genomes was carried out using the JupiterPlot tool (https://github.com/JustinChu/JupiterPlot) which uses minimap2[42] for whole-genome alignments. 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 contigs from the two assemblies being compared are displayed in the Jupiter plots. Whole-genome alignments of *w*Moz, *w*Mpe as well as the *w*Mhie genome against *w*Cle as the reference genome were performed using the nucmer utility in MUMmer4[43] with default parameters. The resulting alignment blocks were visualized as concentric Circos plots[44] drawn using the R package circlize v0.4.10 [45]. The completeness of the protein-coding content of genomes was assessed using the BUSCO pipeline v5.0 beta using the "proteobacteria_odb10" reference dataset[46]. For global sequence comparisons across multiple *Wolbachia* genomes the average nucleotide identity (ANI) scores were calculated using the OrthoANIu tool v1.2[47]. 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[48] 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 v2.4[49], and the results visually represented as UpSet plots[50] using the R package UpSetR

v1.4.0[51]. Insertion sequence (IS) elements were identified via the ISsaga web server[52] 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. Phage-derived genes and regions were annotated by integrating the outputs from the PHASTER web server[53], the PhiSpy v4.2.6 tool[54] (https://github.com/linsalrob/PhiSpy/), and manual curation based on GenBank annotations. Functional annotation of protein-coding genes was carried out using the eggNOG-Mapper[55] 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[56]. Pseudogene loci and the GenBank accession of the ortholog of their potential parent genes were annotated as a part of the NCBI-PGAP pipeline. If a pseudogene from multiple *Wolbachia* shared the same accession for the parent gene ortholog, it was considered to be a shared pseudogene.

Results

Metagenomic assembly and binning retrieves *Wolbachia* genomes from complex samples

All four isolates sequenced here were obtained as microfilariae from *Mansonella*-positive human subjects. The genomic DNA extracted from such clinical samples is unlikely to contain pure DNA originating from the parasite alone, instead each of the samples is expected to yield a complex assortment of mixed-organism DNA. The analysis of raw genome data from these samples therefore required a metagenomic assembly and binning approach to de-convolute parasite, symbiont and contaminants to permit species-specific genome assemblies. Further

complicating correct assembly of *Wolbachia* genomes, is the presence of lateral gene transfer events where symbiont DNA can integrate into the nuclear genome of the host. Appropriate steps were taken to avoid including these HGT regions in the bacterial assemblies.

The assembled metagenome contigs were analyzed using BlobTools[40], which is used to identify clusters of contigs based on their GC content, and read coverage obtained by mapping back input reads to the assembly, overlaid with sequence-identity based taxonomic classification. Partitioning the data in this manner generated distinct blobs corresponding to the desired target *Wolbachia* genomes from the mixed dataset, the genomes of the worm hosts and contaminating microbiota. (Supplementary Figure S1). The *Wolbachia* contigs identified from the BlobTools analysis were extracted as a separate bin of fasta sequences and analyzed further.

Of the two *M. ozzardi* isolates (Moz1 from Brazil and Moz2 from Venezuela), sufficient coverage was obtained from Moz2 that yielded a draft *w*Moz assembly of 1,073,310 bp in size comprised of 93 contigs. The N50 size of this assembly is 17.225 kb and the largest contig is 37.481 kb (Table 1). Of the two *M. perstans* isolates, Mpe1 and Mpe2, only Mpe1 yielded a *Wolbachia* assembly. The *w*Mpe assembly is comprised of 1,058,123 bp in 170 contigs, with a N50 size of 10.041 kb and the largest contig is 28.485 kb (Table 1).

The quality and correctness of these *de novo* assemblies were confirmed by various comparative analyses. Whole genome alignment of *w*Moz and *w*Mpe visualized as a Jupiter plot demonstrates high sequence similarity and co-linearity of these independently assembled, closely related *Wolbachia* (Figure 1). A comparison was also made to the most closely related complete genome available in the Clade F supergroup, namely *w*Cle, where high sequence synteny and co-linearity were also observed (Figures 2A and 2B).

To determine the nature of the gaps in the draft assemblies, the whole genome alignments were investigated in more detail. Interestingly, most of the regions absent from *w*Moz and *w*Mpe assemblies corresponded to regions containing IS element transposons in the *w*Cle genome (Figure 3). In a similar analysis using the recently available draft genome of *w*Mhie, the missing regions were again found to be associated with IS elements in *w*Cle genome (Figure 3). Based on this analysis, it appears that most of the protein coding regions have been assembled into these draft genomes, and only the IS elements remain absent from the current assemblies. Due to their repetitive nature, IS elements are known to present a technical hurdle in assembling complete genomes, leading to fragmented assemblies[57,58].

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.beta) 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), with such scores being typical for *Wolbachia*, even with complete genomes. For comparison, the corresponding scores for other *Wolbachia* were also calculated. The BUSCO scores for *Wolbachia w*Ov from the human filarial parasite *Onchocerca volvulus* and *w*Oo, present in the related bovine parasite *Onchocerca ochengi* are 76.7% and 75.8% respectively, and 79.5% for *w*Bm from the human filarial parasite *Brugia malayi* (Supplementary Figure S2). Within supergroup F, the BUSCO score for filarial *Wolbachia w*Mhie is 79% and is 81.7% for the arthropod *Wolbachia w*Cle.

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

The metagenomic assembly and binning approach also enables the measurement of the relative amounts of *Wolbachia* in different isolates, by comparing the sequencing reads coverage of the *Mansonella* and *Wolbachia* specific contig bins in the metagenomic assemblies (Table 2). In the *M. ozzardi* sample (Moz2), the median coverage of contigs classified as *Mansonella* was observed to be 360X, while that of the *Wolbachia* contigs was found to be 13X. Assuming 1000 nematode cells per microfilaria[59], each parasite is estimated to harbor 37 *Wolbachia* cells. For the other *M. ozzardi* isolate (Moz1) the median coverage for nematode contigs was only 57X and this low coverage is the most likely reason why this isolate did not yield sufficient *Wolbachia* contigs in its metagenomic assembly. Therefore, the 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 contigs was 1032X and the *Wolbachia* coverage was 30X, yielding an estimated titer of 30 *Wolbachia* per microfilaria. In the other *M. perstans* isolate Mpe2, where only 12 *Wolbachia* contigs were produced in the metagenomic assembly, the median coverage for nematode contigs is relatively high 633X, indicating that the lack of *Wolbachia* contigs is not due to insufficient depth of coverage and is rather most likely due to a very low titer of *Wolbachia* in this sample. 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 on detection of *Wolbachia* in *M. perstans*[25–27,31–33].

Genome comparisons with other *Wolbachia*

The sequence similarity and divergence between different *Wolbachia* genomes from filarial nematodes, a plant parasitic nematode, and arthropod hosts representing different supergroups (Supplementary Table S1) was measured using the genome-wide average nucleotide identity (gANI) and digital DNA–DNA hybridization (dDDH) scores. The gANI scores (Figure 4A) for closely related sympatric *Wolbachia* are: 99% for the *w*Oo : *w*Ov pair from *Onchocerca spp*. and 99% for the *w*Bm : *w*Bpa pair from *Brugia spp*. Within supergroup F, high gANI scores were observed as expected; namely 97% between the *w*Moz : *w*Mpe pair, and 92% between *w*Moz : *w*Cle and *w*Mpe : *w*Cle pairs (Figure 4A). *w*Mhie shared a higher similarity with *w*Cle (95%), than to *w*Moz and *w*Mpe (92%). Comparisons between Clade F members and those of other clades generated gANI scores of a maximum value of 85% (Figure 4A).

Using dDDH, a more sensitive metric for sequence divergence between very closely related species, the dDDH score of the *w*Moz : *w*Mpe pair was found to be 71.7% (Figure 4B). For comparison, the dDDH scores for other pairs of closely related *Wolbachia* are: 96.1% for the *w*Oo : *w*Ov pair from *Onchocerca spp*. (Figure 4C) and 93.3% for the *w*Bm : *w*Bpa pair from *Brugia spp*. (Figure 4D). Within supergroup F, the dDDH scores of more diverged pairs, namely *w*Moz : *w*Cle and *w*Mpe : *w*Cle are only ~44%, while the scores obtained for *w*Mhie : *w*Cle is considerably higher at 61.8%.

The core proteome of filarial *w*Mpe and *w*Moz

Orthologs for *w*Moz and *w*Mpe proteins were identified with OrthoFinder[49] using the predicted proteomes of various *Wolbachia*, including complete genomes representing each arthropod *Wolbachia* supergroup, all filarial *Wolbachia* genomes and the only *Wolbachia* genome from a

plant parasitic nematode (Supplementary Table 2). From a total input set of 17,720 proteins, 96% of proteins (n = 17,020) could be clustered into 1,440 orthogroups that were present in at least two *Wolbachia* proteomes, while 1.5% of proteins (n = 253) were grouped into 65 orthogroups that were species-specific (Supplementary Table S2-A). Only 2.5% (n = 447) of all the input proteins could not be assigned to any orthogroups. For *w*Moz and *w*Mpe, 851 and 821 of their protein-coding genes (~95%) were assigned to one of the 1,440 shared orthogroups, while 35 of *w*Moz proteins and 39 of *w*Mpe proteins (less than 5% in both cases) could not be assigned to any orthogroup (Supplementary Table S2-B, C). These trends are very similar for the other genomes analyzed, with more than 95% of proteins for each *Wolbachia* assigned to a shared orthogroup, and the proportion of potentially genome-specific orthogroups being lower than 5% (Supplementary Table S2-C). The only exception to this trend was observed for the *Folsomia candida Wolbachia* (*w*Fol) genome, for which 9.6% of its protein-coding genes were clustered into *w*Fol-specific orthogroups, while 9% of its proteins could not be assigned to any orthogroups.

A total of 411 orthogroups were found to be conserved across all the *Wolbachia* analyzed (Figure 5). Interestingly, 29 orthogroups were found that are present in *w*Mpe and *w*Moz, but not in the two other clade F *Wolbachia*, *w*Cle and *w*Mhie. However, further analysis revealed most of these genes to be hypothetical proteins with a median size of 80 aa. From these patterns of shared orthologs, it can be seen that the set of proteins encoded in supergroup F *Wolbachia* genomes, particularly *w*Moz and *w*Mpe, do not deviate significantly from those of other filarial *Wolbachia*.

Mobile genetic elements in *w*Moz and *w*Mpe genomes

Genes encoding various kinds of transposable elements, including IS elements, were identified based on annotation descriptions obtained from the PGAP pipeline. The *w*Moz genome assembly has 25 transposable elements (3 intact protein coding genes and 22 pseudogenes, Supplementary Table 3A), while the *w*Mpe genome assembly contains 23 transposable elements (1 intact protein coding gene, 22 pseudogenes, Supplementary Table 3B). Applying the same search strategy to the *w*Mhie genome assembly identified 66 transposon elements (16 intact protein coding genes, 50 pseudogenes, Supplementary Table 3C). Within the set of these transposable elements, the *w*Moz assembly harbors only 1 intact IS3 family gene and 12 pseudogenes from various families of IS elements. The *w*Mpe draft genome contains 10 pseudogenes from various IS families but no intact IS element gene, while 14 intact and 42 pseudogenized IS elements were identified in the wMhie genome. Additional search for IS elements in *w*Moz and *w*Mpe genome assemblies using the ISsaga web-server[52] did not reveal any other IS elements in these genomes. In the complete *w*Cle genome, ISsaga web-service could identify 149 IS elements, with the IS5 family being the most abundant (121 copies). The low number of IS elements observed in *w*Moz and *w*Mpe is most likely due to the limitations of short-read data in capturing repeated elements.

Group II introns are a common mobile element found in high copy numbers in many *Wolbachia* genomes, e.g. 53 copies in the *w*AlbB genome[57]. Both the *w*Moz and *w*Mpe assembly contain only 1 pseudogene derived from a Group II intron, while no intact gene or pseudogene for these elements could be found in *w*Mhie and *w*Cle genomes as reported earlier[6].

A search for prophage elements identified 9 phage derived genes in both *w*Moz and *w*Mpe assemblies of which 4 have been pseudogenized in both *Wolbachia* (Supplementary Table

3D). A gene encoding gp6-like head-tail connector phage protein and a pseudogene derived from a phage portal protein were found only in *w*Moz, while a pseudogene derived from major tail-tube phage protein was found only in *w*Mpe. This is in accordance with the trend of none or very few phage derived genes in filarial *Wolbachia* genomes[6].

Analysis of various biological pathways in *w*Moz and *w*Mpe genomes

For a detailed functional analysis, the proteins encoded by the *w*Moz and *w*Mpe genomes were annotated using the eggNOG database (Supplementary Tables S4A-B). Additionally, biosynthetic pathway annotations for the *w*Cle genome were obtained from the KEGG database and corresponding orthologs were identified in *w*Moz and *w*Mpe. The 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 Tables S5A-E). 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)[60] were also present in both *w*Moz and *w*Mpe.

Biotin biosynthesis genes are absent in *w*Mpe and *w*Moz draft assemblies

The bedbug *Wolbachia* *w*Cle genome harbors an operon encoding the enzymes involved in biosynthesis of biotin, supporting a nutritional mutualism between the bedbug and its *Wolbachia*[29]. Interestingly, this operon has most likely been laterally acquired by *w*Cle from a

co-infecting *Rickettsia* or *Cardinium* symbiont[29]. Given the paucity of genome data on other members of supergroup F, it is not clear whether this pathway is also present in other *Wolbachia* from this supergroup. Searches of the *w*Moz and *w*Mpe proteomes did not reveal any members of the biotin pathway suggesting that unlike the *w*Cle and bedbug symbiosis, biotin supplementation is not the basis of mutualism between these *Wolbachia* and their filarial host. The absence of biotin pathway genes is unlikely to be due to the incomplete nature of the draft assemblies. Given the high BUSCO score (~80%) of the assemblies, there is a high probability that at least fragments of a few of the 6 genes in the biotin pathway would have been detected if present in the genomes. Further, the biotin pathway genes were also found to be absent in the draft genome of *w*Mhie - the other filarial *Wolbachia* from the supergroup F[6]. In contrast, the biosynthetic genes of other vitamin pathways present in *w*Cle were easily found in *w*Mpe and *w*Moz (Supplementary Table S5A-E).

Shared pseudogenes point to a common ancestor of *w*Moz and *w*Mpe being present in the common ancestor of *M. perstans* and *M. ozzardi*

Intracellular bacteria such as *Wolbachia* are expected to accumulate deleterious mutations over evolutionary timescales, with one consequence being the formation of pseudogenes[61]. Analysis of these pseudogenes can also shed light on the nature of genes that are dispensable during a symbiotic association with the host organism. Pseudogenes in *w*Moz (n = 118) and *w*Mpe (n = 132) were annotated as a part of the NCBI PGAP pipeline and compared with corresponding annotations in *w*Mhie (n = 87) and *w*Cle (n = 212). These annotations include information on the ortholog of ancestral protein-coding gene for each pseudogene (Supplementary Table 6). If the

same ancestral gene was found to have converted to a pseudogene in two or more *Wolbachia*, it was counted as a shared pseudogene (Figure 6). Transposons such as IS elements were excluded from this analysis due to their tendency to frequently pseudogenize.

For each *Wolbachia*, the majority of its pseudogenes were found to be unique to its genome, and only 1 pseudogene was shared among all the 4 *Wolbachia*. In pairwise comparisons of *Wolbachia*, only 2 to 7 pseudogenes were common among pairs of *Wolbachia*, except for the *w*Moz : *w*Mpe pair, which share a much larger set (n = 38). This large proportion of common pseudogenes suggests that these pseudogenes were already present in the last common ancestor (LCA) of *w*Moz and *w*Mpe.

Discussion

To date, there are only two complete *Wolbachia* genomes available from human filarial parasites; *w*Bm from *Brugia malayi*[62], one of several parasites responsible for lymphatic filariasis or elephantiasis, and *w*Ov from *O. volvulus*, the causative agent of onchocerciasis or river blindness[63]. These genomes belong to different phylogenetic lineages - *w*Bm is a member of supergroup D, whereas *w*Ov has been assigned to supergroup C. This is the first report describing the genomes of supergroup F *Wolbachia* from *Mansonella*. In this study a metagenomic approach was utilized to successfully retrieve the *w*Moz genome from one isolate of *M. ozzardi* collected in Venezuela and the *w*Mpe genome from one isolate of *M. perstans* from Cameroon. The metagenomic assembly and binning approach that combines information from GC content of assembled contigs in conjunction with their read-coverage played a critical role in the generation of high-quality, draft genomes for *w*Mpe and *w*Moz, from complex clinical

samples containing a mixture of DNA from the human host, the filarial parasite and its intracellular symbiont and possibly other organisms that might be present in humans. The use of BlobTools[40] for binning of contigs has been successfully employed previously to disentangle genomes of the *Wolbachia w*Dimm from its host *Dirofilaria immitis*[64,65], a filarial parasite of dogs and cats that occasionally infects humans.

The metagenomic approach also enables an accurate determination of levels of endosymbiont present in the worm host. In a previous analysis of *Wolbachia* burden in different strains and stages of *D. immitis*, use of a BlobTools-like approach identified distinct blobs of contigs with levels of coverage consistent with high or low *Wolbachia* abundance[65]. Performing a similar analysis for *Mansonella* samples in the current study, the number of *Wolbachia* was estimated to be around 30 cells per microfilaria for samples Moz2 and Mpe1. This level of *Wolbachia* in *Mansonella* is considerably lower than the reported levels of 100 *Wolbachia* per microfilaria in *Brugia malayi*[66]. For the other *M. perstans* sample Mpe2 obtained from the same geographical location as Mpe1, only 12 small contigs corresponding to *Wolbachia* were found at very low coverage, demonstrating that the endosymbiont was present at extremely low levels, precluding accurate titer estimation. These observations indicate that even within a locality, *Wolbachia* levels can vary between different isolates, and could be a possible explanation for conflicting reports on the presence and absence of *Wolbachia* in *M. perstans*[25–27,31–33]. In addition to potential natural variation, previous exposure to antibiotics can also be a cause of very low *Wolbachia* levels, and could be another explanation for the occasional failures to detect *Wolbachia* in PCR-based assays[25–27,31–33].

The observed differences in *Wolbachia* detection have raised doubts on the essentiality of *Wolbachia* in *M. perstans*[32]. However, the success of antibiotic treatments in reducing *M.*

perstans microfilaremia supports an essential role of *Wolbachia* in this species[25–27]. Interestingly, *Wolbachia* has recently been reported in *Mansonella sp.* "DEUX" [33], a potentially new species of *Mansonella* parasite of humans, which is closely related to *M. perstans*[67]. To date, *Wolbachia* has been consistently found in all *M. ozzardi* isolates[5,28], including the two isolates in the current study. Therefore, *Wolbachia* is most likely essential and present in all *Mansonella* parasites of humans, but can be present at different levels, sometimes below the limit of detection in *M. perstans*.

Analysis of whole-genome alignments of the *w*Mpe and *w*Moz to the complete *w*Cle genome in the same supergroup F identified regions of high sequence similarity and synteny, and also revealed that the location of alignment gaps mainly corresponded to the locations of IS elements in *w*Cle genome. Due to their repetitive nature, the high-copy number IS elements are challenging to assemble with short-read data alone, and produce assemblies fragmented into multiple contigs, with the contig breakpoints mostly caused by the un-assembled IS elements[57,58]. Based on these observations, it is quite likely that both *w*Moz and *w*Mpe contain some as yet unidentified IS elements present in high copy numbers in their genomes. Consistent with this, 10 pseudogenized IS elements could be identified in both *w*Moz and *w*Mpe assemblies and only one 1 intact IS element could be identified in *w*Moz. The pseudogene copies diverge in sequence from their parent gene and therefore become low-copy variants that can be successfully assembled even with short-read data. Assuming at least one IS element in each of the alignment gaps, the number of IS elements in *w*Moz and *w*Mpe is predicted to be similar to the number of their contigs, that is around 100 copies. Using the same annotation pipeline on the *w*Mhie draft genome, 42 pseudogenized and 14 intact IS elements could be identified. Together, these observations point to presence of around 50 to 100 IS elements in filarial *Wolbachia* from

supergroup F. This is in contrast to the trend in supergroups C and D, where fewer than 5 IS elements could be identified in complete genomes[6,57,68]. It is common for arthropod *Wolbachia* from other supergroups (such as A and B) to have hundreds of copies of IS elements, for example, comprising as high as 13% of the genome sequence in *w*AlbB[57]. Supergroup F is unique in having members from both filarial and arthropod hosts. It is interesting that the number of IS elements in filarial *Wolbachia* from this clade is intermediate between those observed in other arthropod and filarial *Wolbachia*.

The *w*Moz and *w*Mpe genomes enable analysis of their evolutionary history from multiple perspectives. Insights can be gained from comparison of pseudogene content across different supergroup F *Wolbachia*. Since pseudogenes are evolutionary accidents, and the probability of such disruptions happening in the same gene in two independent lineages is low, the presence of multiple shared pseudogenes suggests that these pseudogenes were already present in the last common ancestor (LCA) of *w*Moz and *w*Mpe. Since these endosymbionts are restricted to survive only within their host nematodes, it can be extrapolated that the LCA of *w*Moz and *w*Mpe should have been already present in the last common ancestor of *M. ozzardi* and *M. perstans*. After the split, both *w*Moz and *w*Mpe have continued to accumulate lineage-specific pseudogenes (n = 56 for *w*Moz, n = 62 for *w*Mpe).

Comparisons of genome-wide sequence similarity within supergroup F were also made using the dDDH score. The dDDH scores simulate the results of DNA : DNA hybridization, and provide a sensitive measure of sequence similarity and divergence between closely related bacteria at the whole-genome level. Additionally, this measure has been used to determine species boundaries, with dDDH scores above 70% typically indicating the same species[48]. The dDDH score for the *w*Moz : *w*Mpe pair was found to be 71.7%, while the corresponding scores

between other closely related but distinct *Wolbachia* e.g., *w*Bm : *w*Bpa pair and *w*Oo : *w*Ov pair were even higher (93.3% and 96.1% respectively). Since each of these *Wolbachia* pairs (e.g., *w*Moz and *w*Mpe) are found in distinct nematode hosts and cannot exchange genetic material, they are biologically distinct species. Therefore, the threshold used to delineate species boundaries in free-living bacteria is most likely not directly applicable to intracellular endosymbionts, which would have very different population dynamics as compared to free-living species. The dDDH score is nonetheless a useful metric for measuring sequence divergence. The lower dDDH scores of the *w*Moz : *w*Mpe pair as compared to the *w*Bm : *w*Bpa and *w*Ov : *w*Oo pairs indicates that the nucleotide sequence divergence between *Wolbachia* of *Mansonella* is much higher than the corresponding divergence in *Wolbachia* from either the *Brugia* or *Onchocerca* pair. While *w*Moz and *w*Mpe are both derived from filarial parasites of humans, they still show higher divergence than other pairs that comprise of one *Wolbachia* from a human filarial parasite (*w*Bm, *w*Ov) and one from a non-human filaria (*w*Bpa, *w*Oo). One potential cause for the higher divergence observed in *w*Moz : *w*Mpe pair could be the *Mansonella* lineages splitting into two species (*M. perstans* and *M. ozzardi*) much earlier than the splitting within *Brugia* or *Onchocerca* lineages. Additionally, other ecological factors such as geographical location and vectors involved might contribute to sequence divergence of filarial nematodes and their endosymbionts.

Orthology analysis and annotation of key biological pathways were performed to infer biological functions of the *w*Moz and *w*Mpe endosymbionts. These comparisons included *Wolbachia* genomes from other filarial hosts, as well as representative genomes from other supergroups. Similar to other *Wolbachia*, genes encoding enzymes of various biosynthetic pathways, namely heme, purine, pyrimidine and riboflavin were present in *w*Moz and *w*Mpe.

The proposed anti-*Wolbachia* drug target PPDK involved in the glycolysis/gluconeogenesis pathway[60] was conserved in all *Wolbachia*. Doxycycline, a commonly used anti-*Wolbachia* antibiotic that targets protein synthesis machinery, is reported to be effective against *Mansonella perstans*[25–27]. Consistent with this, all ribosomal subunits were found in *w*Moz and *w*Mpe.

The supergroup F *Wolbachia w*Cle provides biotin as a nutritional supplement to its host bedbug and contains the complete pathway for biotin biosynthesis[29]. However, none of the genes of the biotin pathway could be found in *w*Moz, *w*Mpe or in *w*Mhie, indicating that biotin supplementation is not the basis of their filarial symbiosis. The biotin pathway in the *w*Cle genome has been acquired via LGT from another *Rickettsia*[29] and does not seem to be a feature common to other *Wolbachia* of supergroup F.

In summary, high quality draft genomes of *Wolbachia w*Moz and *w*Mpe from human filarial parasites *M. ozzardi* and *M. perstans* have been generated from genomic analysis of multiple isolates from different continents. These genomes provide an important resource for studies of symbiosis, evolution, comparative genomics, as well as searches for new drug targets. In addition, the successful use of a metagenomic assembly and binning approach for obtaining endosymbiont genomes exemplifies the wider applicability of this method to other systems where the DNA samples can only be obtained as complex mixtures of multiple organisms. Since the genomes from the host and endosymbiont are sequenced simultaneously, this approach uniquely enables a method for directly and robustly estimating the relative abundance of different organisms in a metagenomic sample.

Acknowledgments

The authors thank Jeremy Foster 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 wMoz and wMpe, and other Wolbachia from supergroup F

	***w*Moz**	***w*Mpe**	***w*Mhie**	***w*Cle**
Assembled genome size	1,073,310	1,058,123	1,025,329	1,250,060
Number of contigs	93	170	208	1
Largest contig	37.481 kb	28.485 kb	21.863 kb	-
Contig N50 size	17.225 kb	10.041 kb	6.845 kb	-
%GC	35.61	35.45	36.1	36.3
Predicted genes	1,079	1,058	1,081	1,238
Predicted proteins	888	864	907	1,020
Predicted pseudogenes	145	153	137	177
tRNA genes	36	34	30	34
BUSCO score (%)	78.1	75.3	79.0	81.7

Table 2. Estimation of *Wolbachia* levels per microfilaria

Host species	**Isolate name, Location**	**Median coverage of Mansonella contigs (M)**	**Number and total size of assembled *Wolbachia* contigs**	***Wolbachia* genome assembled**	**Median coverage of *Wolbachia* contigs (W)**	**Number of *Wolbachia* per microfilaria* (1000 x W/M)**
M. ozzardi	Moz1 , Brazil	57	173 contigs, 0.12 Mb	No	3	NA[#]
M. ozzardi	Moz2, Venezuela	360	93 contigs, 1.073 Mb	*w*Moz	13	37

M. perstans	Mpe1, Cameroon	1,032	170 contigs, 1.058 Mb	*w*Mpe	30	30
M. perstans	Mpe2, Cameroon	633	5 contigs, 8.781 kb	No	NA[#]	NA[#]

***** Assuming 1000 cells per microfilaria

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

Figure Legends

Figure 1. Jupiter plot showing high 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 contigs are displayed as colored boxes in the left semi-circle, while the *w*Mpe contigs are shown as gray boxes in the right semi-circle. The syntenic regions are marked by bands connecting contigs 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.

Figure 2. Jupiter plots showing synteny between (A) *w*Moz and (B) *w*Mpe with 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 contigs 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.

Figure 3. Comparison of genome sequences of all supergroup F *Wolbachia*

Whole genome alignments of *w*Moz (blue ring), *w*Mpe (light-blue ring) and *w*Mhie (innermost gray ring) to the *w*Cle chromosome (outermost black circle) are visualized as a circos plot. The red bars mark the locations of the IS elements in the *w*Cle genome.

Figure 4. 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, and 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.

Figure 5. Orthogroups conserved in *w*Mpe and *w*Moz compared to 18 other *Wolbachia* genomes

Orthology relationships between proteins encoded by *w*Moz and *w*Mpe and 18 other *Wolbachia* were determined using OrthoFinder and visualized as UpSet plots. Black dots represent presence of an orthogroup while light gray dots denote absence. The *Wolbachia* supergroup is indicated except for the currently un-assigned *Wolbachia* from *Ctenocephalides felis* J (*w*CfeJ) and *Ctenocephalides felis* T (*w*CfeT).

Figure 6. Venn diagram comparing pseudogenes in supergroup F *Wolbachia*

Pseudogenes for *w*Moz and *w*Mpe were annotated as a part of the NCBI PGAP pipeline. Transposons such as IS elements were excluded from this analysis due to their tendency to frequently convert into pseudogenes.

Supplementary Information

Figure S1. BlobTools based binning of the metagenomes assembled from *Mansonella* isolates (A) *M. ozzardi* isolate Moz1, from Brazil (B) *M. ozzardi* isolate Moz2, from Venezuela (C) *M. perstans* isolate Mpe1, from Cameroon and (D) *M. perstans* isolate Mpe2, from Cameroon.

Each dot in the plot represents a contig 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 contig. 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* contigs, but very few or very small contigs were actually recovered for these samples.

Figure S2. BUSCO scores of various *Wolbachia*

The BUSCO pipeline was used to measure the proportion of highly conserved, single copy orthologs (BUSCO groups). The set of reference BUSCO groups was set to the lineage "proteobacteria_odb10", which contains 219 BUSCOs derived from proteobacterial species. The *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.

Table S1. *Wolbachia* genomes used for comparative analysis with *w*Moz and *w*Mpe genomes

Table S2: Orthology analysis using OrthoFinder

Table S2-A. Orthogroups and their member genes for different *Wolbachia*

Table S2-B. Genes not assigned to any orthogroups.

Table S2-C. Overall OrthoFinder statistics for each *Wolbachia* genome analyzed.

Table S3: Mobile genetic elements in *w*Moz and *w*Mpe genome assemblies. (A) Transposons and Group II introns in *w*Moz (B) Transposons and Group II introns in *w*Mpe (C) Transposons in *w*Mhie (D) Phage derived genes in *w*Moz and *w*Mpe

Table S4: eggNOG annotations of (A) wMoz and (B) wMpe proteins

Table S5: Various biological pathways in *w*Moz and *w*Mpe and other supergroup F *Wolbachia*. (A) Heme pathway (B) Purine (C) Pyrimidine (D) Type IV and Type VI secretion systems

Table S6: Pseudogenes present in the genomes of 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. Lefoulon, E. *et al.* Breakdown of coevolution between symbiotic bacteria *Wolbachia* and their filarial hosts. *PeerJ* **4**, e1840 (2016).
6. 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).
7. Lefoulon, E. *et al.* Pseudoscorpion *Wolbachia* symbionts: diversity and evidence for a new supergroup S. *BMC Microbiol.* **20**, 188 (2020).
8. Downes, B. & Jacobsen, K. A systematic review of the epidemiology of mansonelliasis. *Afr. J. Infect. Dis.* **4**, (2010).
9. Simonsen, P. E., Onapa, A. W. & Asio, S. M. *Mansonella perstans* filariasis in Africa. *Acta Trop.* **120**, S109–S120 (2011).
10. 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).

656 11. Mediannikov, O. & Ranque, S. Mansonellosis, the most neglected human filariasis. *New Microbes*
657 *New Infect.* **26**, S19–S22 (2018).
658 12. Ta-Tang, T.-H., Crainey, J. L., Post, R. J., Luz, S. L. & Rubio, J. M. Mansonellosis: current perspectives.
659 *Res. Rep. Trop. Med.* **9**, 9–24 (2018).
660 13. Bélard, S. & Gehringer, C. High Prevalence of *Mansonella* Species and Parasitic Coinfections in
661 Gabon Calls for an End to the Neglect of *Mansonella* Research. *J. Infect. Dis.* **223**, 187–188 (2021).
662 14. Tavares da Silva, L. B. *et al.* Molecular Verification of New World *Mansonella perstans* Parasitemias.
663 *Emerg. Infect. Dis.* **23**, 545–547 (2017).
664 15. Raccurt, C. P. *Mansonella ozzardi* and its vectors in the New World: an update with emphasis on the
665 current situation in Haiti. *J. Helminthol.* **92**, 655–661 (2018).
666 16. Calvopina, M., Chiluisa-Guacho, C., Toapanta, A., Fonseca, D. & Villacres, I. High Prevalence of
667 *Mansonella ozzardi* Infection in the Amazon Region, Ecuador. **25**, (2019).
668 17. Kozek, W. J., Palma, G., Henao, A., García, H. & Hoyos, M. Filariasis in Colombia: Prevalence and
669 Distribution of *Mansonella Ozzardi* and *Mansonella (=Dipetalonema) Perstans* Infections in the
670 Comisaría Del Guainía. *Am. J. Trop. Med. Hyg.* **32**, 379–384 (1983).
671 18. Crainey, J. L. *et al.* Deep-sequencing reveals occult mansonellosis co-infections in residents from the
672 Brazilian Amazon village of São Gabriel da Cachoeira. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.*
673 (2020) doi:10.1093/cid/ciaa082.
674 19. Langworthy, N. G. *et al.* Macrofilaricidal activity of tetracycline against the filarial nematode
675 *Onchocerca ochengi*: elimination of *Wolbachia* precedes worm death and suggests a dependent
676 relationship. *Proc. R. Soc. Lond. B Biol. Sci.* **267**, 1063–1069 (2000).
677 20. Bazzocchi, C. *et al.* Combined ivermectin and doxycycline treatment has microfilaricidal and
678 adulticidal activity against *Dirofilaria immitis* in experimentally infected dogs. *Int. J. Parasitol.* **38**,
679 1401–1410 (2008).

680 21. Taylor, M. J. *et al.* Preclinical development of an oral anti-*Wolbachia* macrolide drug for the
681 treatment of lymphatic filariasis and onchocerciasis. *Sci. Transl. Med.* **11**, (2019).
682 22. Clare, R. H. *et al.* Industrial scale high-throughput screening delivers multiple fast acting
683 macrofilaricides. *Nat. Commun.* **10**, 11 (2019).
684 23. Hübner, M. P. *et al.* In vivo kinetics of *Wolbachia* depletion by ABBV-4083 in *L. sigmodontis* adult
685 worms and microfilariae. *PLoS Negl. Trop. Dis.* **13**, e0007636 (2019).
686 24. Hong, W. D. *et al.* AWZ1066S, a highly specific anti-*Wolbachia* drug candidate for a short-course
687 treatment of filariasis. *Proc. Natl. Acad. Sci.* **116**, 1414–1419 (2019).
688 25. Keiser, P. B. *et al.* Molecular identification of *Wolbachia* from the filarial nematode *Mansonella*
689 *perstans*. *Mol. Biochem. Parasitol.* **160**, 123–128 (2008).
690 26. Coulibaly, Y. I. *et al.* A Randomized Trial of Doxycycline for *Mansonella perstans* Infection. *N. Engl. J.*
691 *Med.* **361**, 1448–1458 (2009).
692 27. Debrah, L. B. *et al.* The Efficacy of Doxycycline Treatment on *Mansonella perstans* Infection: An
693 Open-Label, Randomized Trial in Ghana. *Am. J. Trop. Med. Hyg.* **101**, 84–92 (2019).
694 28. Casiraghi, M., Favia, G., Cancrini, G., Bartoloni, A. & Bandi, C. Molecular identification of *Wolbachia*
695 from the filarial nematode *Mansonella ozzardi*. *Parasitol. Res.* **87**, 417–420 (2001).
696 29. Nikoh, N. *et al.* Evolutionary origin of insect–*Wolbachia* nutritional mutualism. *Proc. Natl. Acad. Sci.*
697 **111**, 10257–10262 (2014).
698 30. Hosokawa, T., Koga, R., Kikuchi, Y., Meng, X.-Y. & Fukatsu, T. *Wolbachia* as a bacteriocyte-associated
699 nutritional mutualist. *Proc. Natl. Acad. Sci.* **107**, 769–774 (2010).
700 31. Grobusch, M. P., Kombila, M., Autenrieth, I., Mehlhorn, H. & Kremsner, P. G. No evidence of
701 *Wolbachia* endosymbiosis with *Loa loa* and *Mansonella perstans*. *Parasitol. Res.* **90**, 405–408 (2003).
702 32. Gehringer, C. *et al.* Molecular Evidence of *Wolbachia* Endosymbiosis in *Mansonella perstans* in
703 Gabon, Central Africa. *J. Infect. Dis.* **210**, 1633–1638 (2014).

704 33. Sandri, T. L. *et al.* Molecular Epidemiology of *Mansonella* Species in Gabon. *J. Infect. Dis.* **223**, 287–
705 296 (2021).
706 34. Basano, S. de A. *et al.* Phase III Clinical Trial to Evaluate Ivermectin in the Reduction of *Mansonella*
707 *ozzardi* infection in the Brazilian Amazon. *Am. J. Trop. Med. Hyg.* **98**, 786–790 (2018).
708 35. Lima, N. F., Gonçalves-Lopes, R. M., Kruize, Y. C. M., Yazdanbakhsh, M. & Ferreira, M. U. CD39 and
709 immune regulation in a chronic helminth infection: The puzzling case of *Mansonella ozzardi*. *PLoS*
710 *Negl. Trop. Dis.* **12**, e0006327 (2018).
711 36. Ritter, M. *et al. Mansonella perstans* microfilaremic individuals are characterized by enhanced type
712 2 helper T and regulatory T and B cell subsets and dampened systemic innate and adaptive immune
713 responses. *PLoS Negl. Trop. Dis.* **12**, e0006184 (2018).
714 37. Poole, C. B. *et al.* In Silico Identification of Novel Biomarkers and Development of New Rapid
715 Diagnostic Tests for the Filarial Parasites *Mansonella perstans* and *Mansonella ozzardi*. *Sci. Rep.* **9**,
716 10275 (2019).
717 38. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic
718 assembler. *Genome Res.* **27**, 824–834 (2017).
719 39. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359
720 (2012).
721 40. Laetsch, D. R. & Blaxter, M. L. BlobTools: Interrogation of genome assemblies. *F1000Research* **6**,
722 1287 (2017).
723 41. Tatusova, T. *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624
724 (2016).
725 42. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
726 43. Marçais, G. *et al.* MUMmer4: A fast and versatile genome alignment system. *PLOS Comput. Biol.* **14**,
727 e1005944 (2018).

728 44. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**,
729 1639–1645 (2009).
730 45. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular
731 visualization in R. *Bioinforma. Oxf. Engl.* **30**, 2811–2812 (2014).
732 46. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing
733 genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**,
734 3210–3212 (2015).
735 47. Yoon, S.-H., Ha, S., Lim, J., Kwon, S. & Chun, J. A large-scale evaluation of algorithms to calculate
736 average nucleotide identity. *Antonie Van Leeuwenhoek* **110**, 1281–1286 (2017).
737 48. Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P. & Göker, M. Genome sequence-based species
738 delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**, 60
739 (2013).
740 49. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics.
741 *Genome Biol.* **20**, 238 (2019).
742 50. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. UpSet: Visualization of Intersecting
743 Sets. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983–1992 (2014).
744 51. Conway, J. R., Lex, A., Gehlenborg, N. & Hancock, J. UpSetR: an R package for the visualization of
745 intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).
746 52. Varani, A. M., Siguier, P., Gourbeyre, E., Charneau, V. & Chandler, M. ISsaga is an ensemble of web-
747 based methods for high throughput identification and semi-automatic annotation of insertion
748 sequences in prokaryotic genomes. *Genome Biol.* **12**, R30 (2011).
749 53. Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.*
750 **44**, W16–W21 (2016).

751 54. Akhter, S., Aziz, R. K. & Edwards, R. A. PhiSpy: a novel algorithm for finding prophages in bacterial
752 genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.* **40**, e126–
753 e126 (2012).
754 55. Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by
755 eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
756 56. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on
757 genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
758 57. Sinha, A., Li, Z., Sun, L. & Carlow, C. K. S. Complete Genome Sequence of the *Wolbachia w*AlbB
759 Endosymbiont of Aedes albopictus. *Genome Biol. Evol.* **11**, 706–720 (2019).
760 58. Lefoulon, E. *et al.* Large Enriched Fragment Targeted Sequencing (LEFT-SEQ) Applied to Capture of
761 *Wolbachia* Genomes. *Sci. Rep.* **9**, 1–10 (2019).
762 59. Basyoni, M. M. A. & Rizk, E. M. A. Nematodes ultrastructure: complex systems and processes. *J.*
763 *Parasit. Dis.* **40**, 1130–1140 (2016).
764 60. Raverdy, S., Foster, J. M., Roopenian, E. & Carlow, C. K. S. The *Wolbachia* endosymbiont of *Brugia*
765 *malayi* has an active pyruvate phosphate dikinase. *Mol. Biochem. Parasitol.* **160**, 163–166 (2008).
766 61. Andersson, S. G. E. & Kurland, C. G. Reductive evolution of resident genomes. *Trends Microbiol.* **6**,
767 263–268 (1998).
768 62. Foster, J. *et al.* The *Wolbachia* Genome of *Brugia malayi*: Endosymbiont Evolution within a Human
769 Pathogenic Nematode. *PLOS Biol.* **3**, e121 (2005).
770 63. Cotton, J. A. *et al.* The genome of *Onchocerca volvulus*, agent of river blindness. *Nat. Microbiol.* **2**,
771 16216 (2016).
772 64. Kumar, S. & Blaxter, M. L. Simultaneous genome sequencing of symbionts and their hosts. *Symbiosis*
773 **55**, 119–126 (2011).

774 65. Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M. & Blaxter, M. Blobology: exploring raw genome
775 data for contaminants, symbionts, and parasites using taxon-annotated GC-coverage plots.
776 *Bioinforma. Comput. Biol.* **4**, 237 (2013).
777 66. McGarry, H. F., Egerton, G. L. & Taylor, M. J. Population dynamics of *Wolbachia* bacterial
778 endosymbionts in *Brugia malayi*. *Mol. Biochem. Parasitol.* **135**, 57–67 (2004).
779 67. Mourembou, G. *et al. Mansonella*, including a Potential New Species, as Common Parasites in
780 Children in Gabon. *PLoS Negl. Trop. Dis.* **9**, e0004155 (2015).
781 68. Cerveau, N., Leclercq, S., Leroy, E., Bouchon, D. & Cordaux, R. Short- and Long-term Evolutionary
782 Dynamics of Bacterial Insertion Sequences: Insights from *Wolbachia* Endosymbionts. *Genome Biol.*
783 *Evol.* **3**, 1175–1186 (2011).
784
785
786

Figure 1 : Jupiter plot showing high synteny between the de novo assembled *w*Moz and *w*Mpe genomes

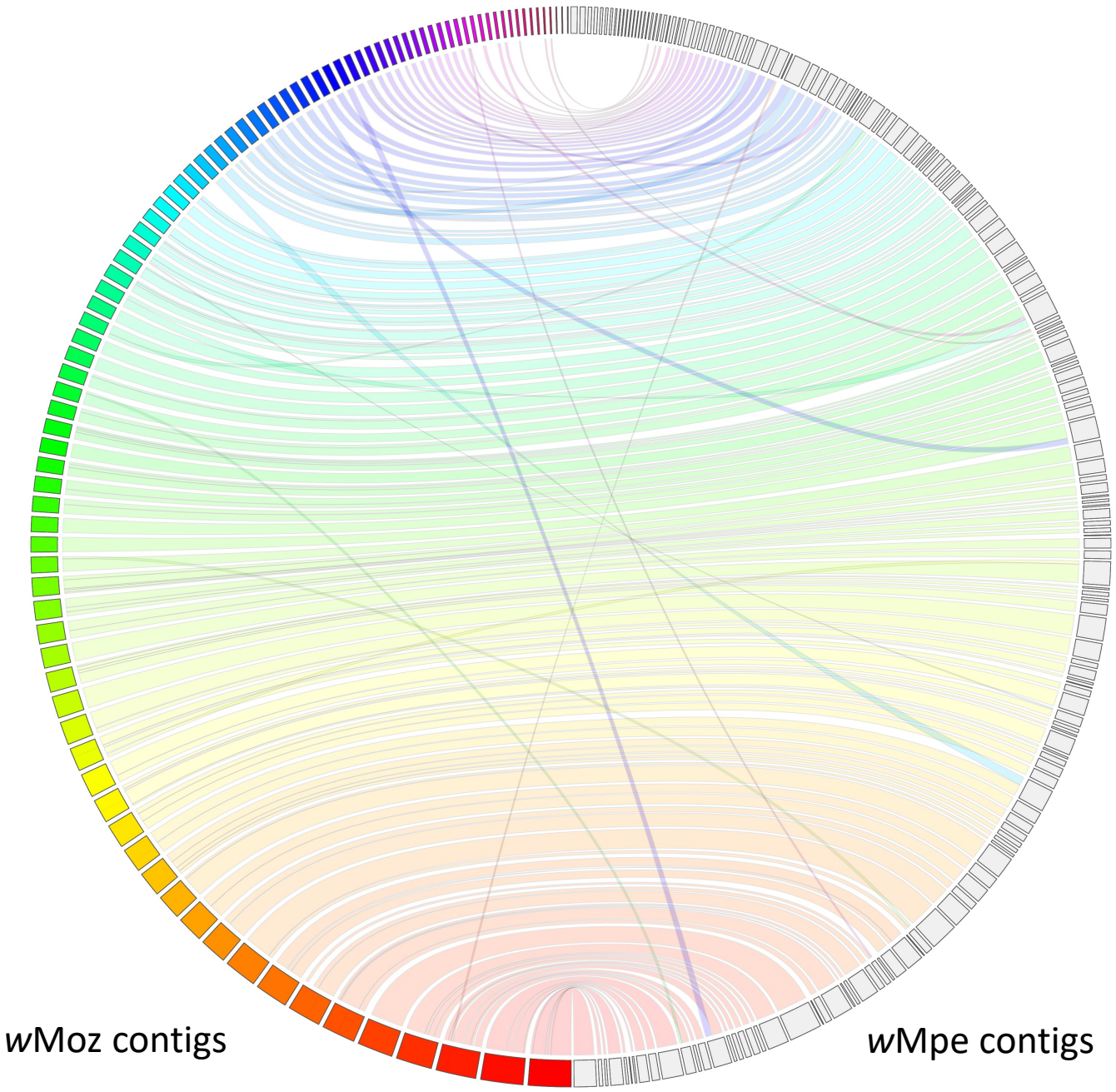
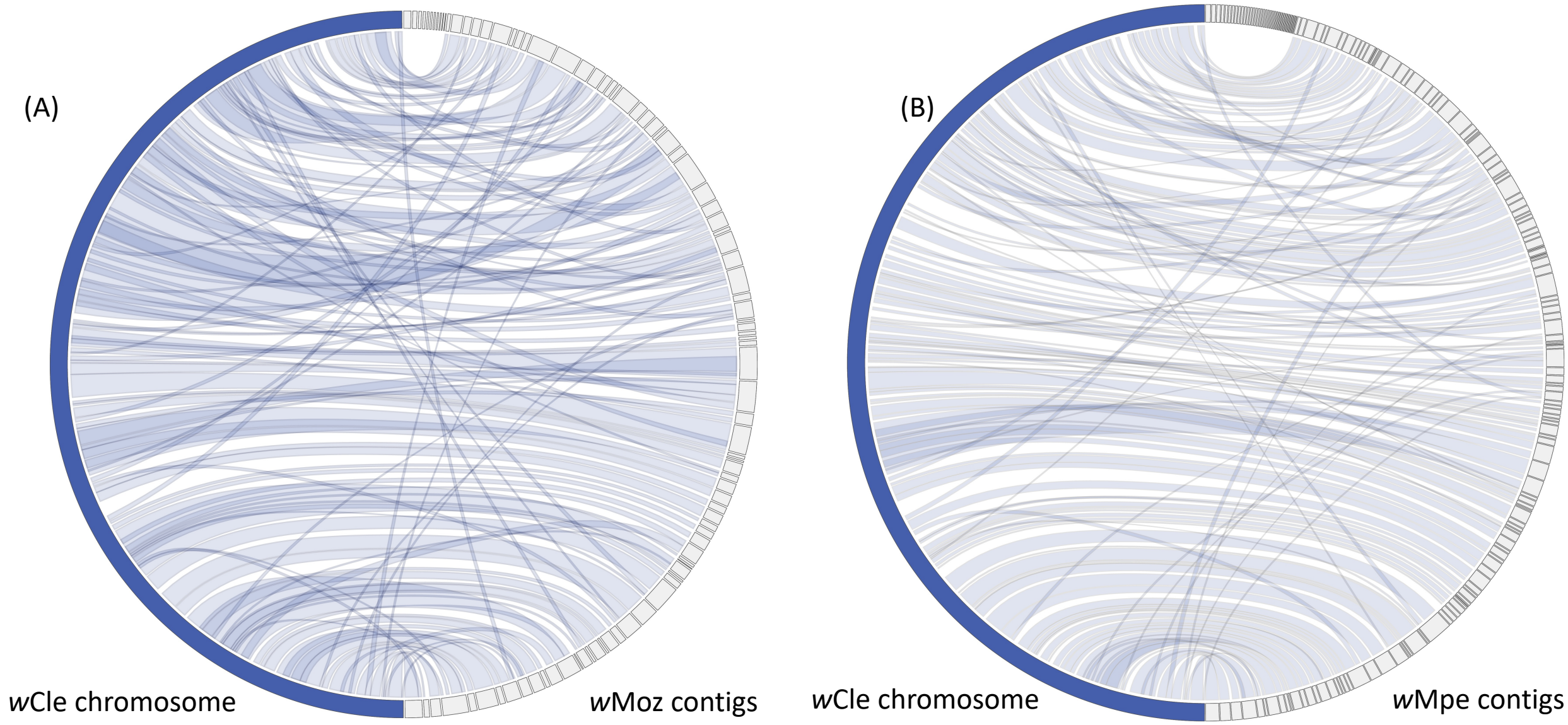


Figure 2 : Jupiter plots showing synteny between (A) *w*Moz and (B) *w*Mpe with the wCle genome



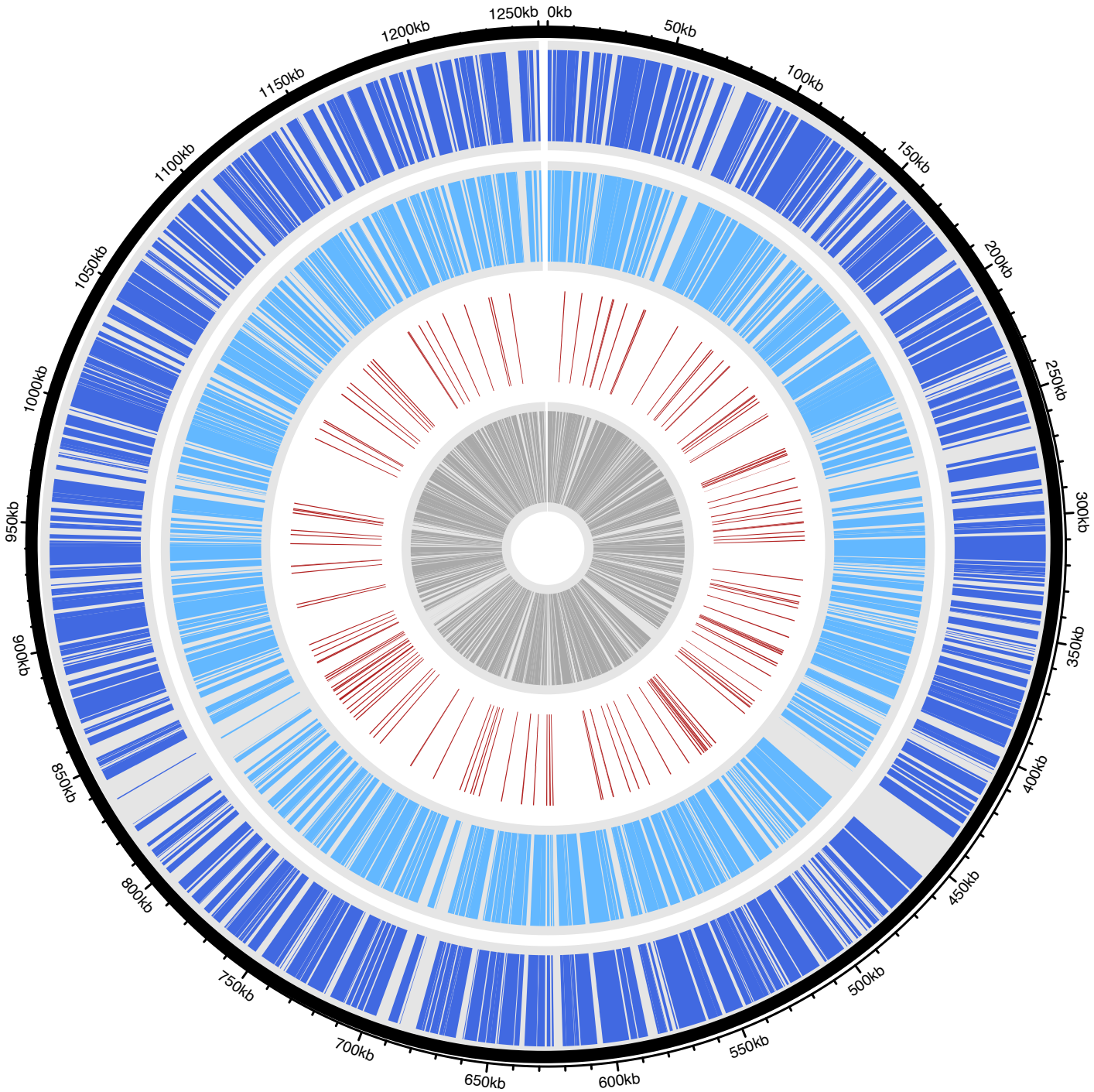


Figure 3
Comparison of genome sequences of all supergroup F *Wolbachia*

Whole genome alignments of *w*Moz (blue ring), *w*Mpe (light-blue ring) and *w*Mhie (innermost gray ring) to the *w*Cle chromosome (outermost black circle) are visualized as a Circos plot. The red bars mark the locations of the IS elements in the *w*Cle genome.

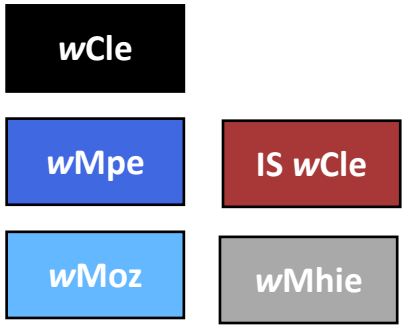
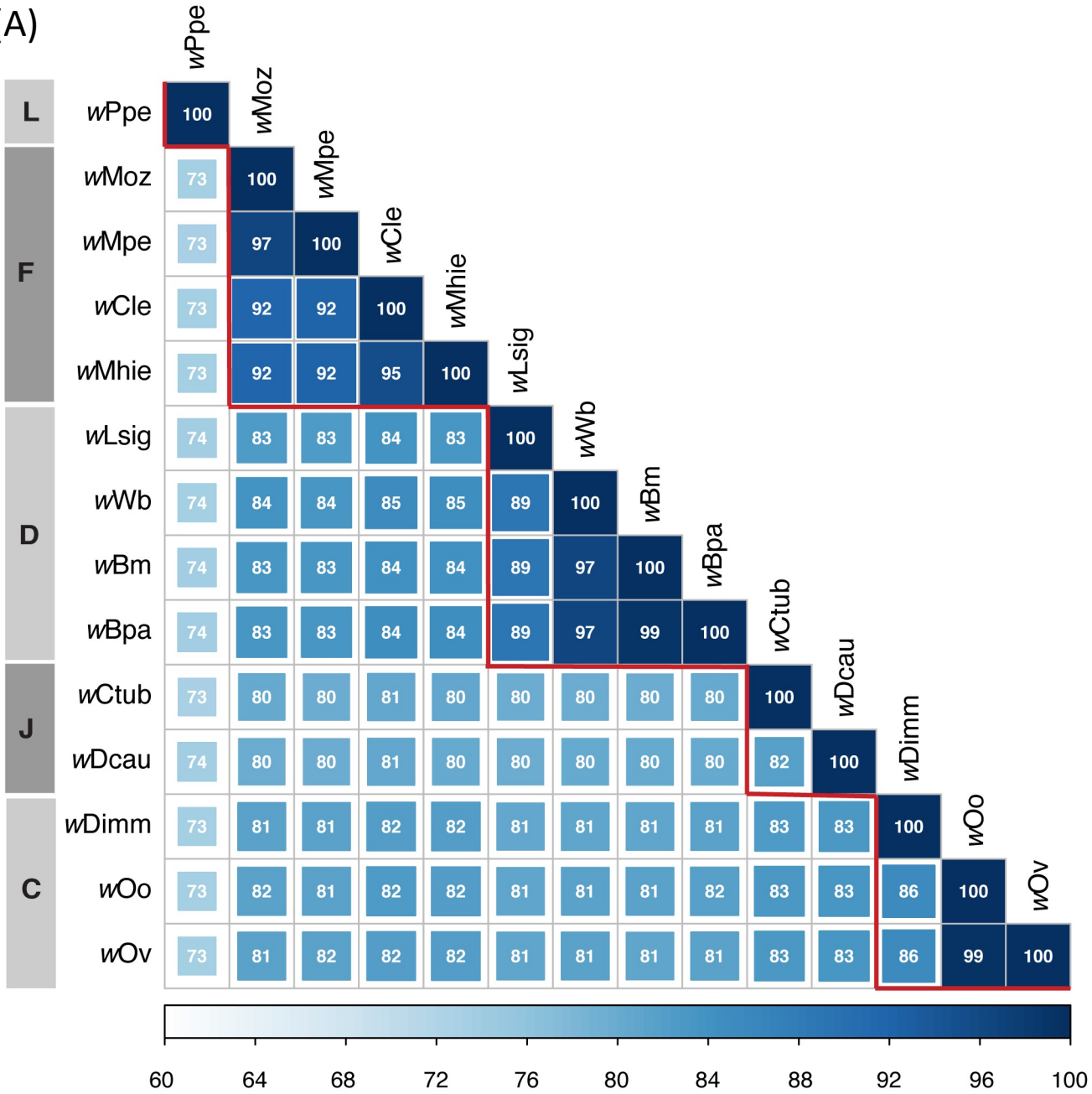


Figure 4: Global comparisons between multiple *Wolbachia* genomes using genome-wide Average Nucleotide Identity (gANI) scores, and digital DNA:DNA Hybridization (dDDH) scores

(A)



(B)

	wMpe	wMoz	wMhie
wMpe			
wMoz	**71.7 [68.7 - 74.5]**		
wMhie	45.7 [43.1 - 48.2]	45.3 [42.7 - 47.9]	
wCle	44.1 [41.5 - 46.6]	43.6 [41.1 - 46.2]	61.80 [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]	70.30 [67.3 - 73.1]	
wLsig	36.5 [34 - 39]	36.40 [33.9 - 38.9]	36.80 [34.3 - 39.3]

Figure 5 :Orthogroups conserved across *w*Mpe, *w*Moz compared to 18 other *Wolbachia* genomes

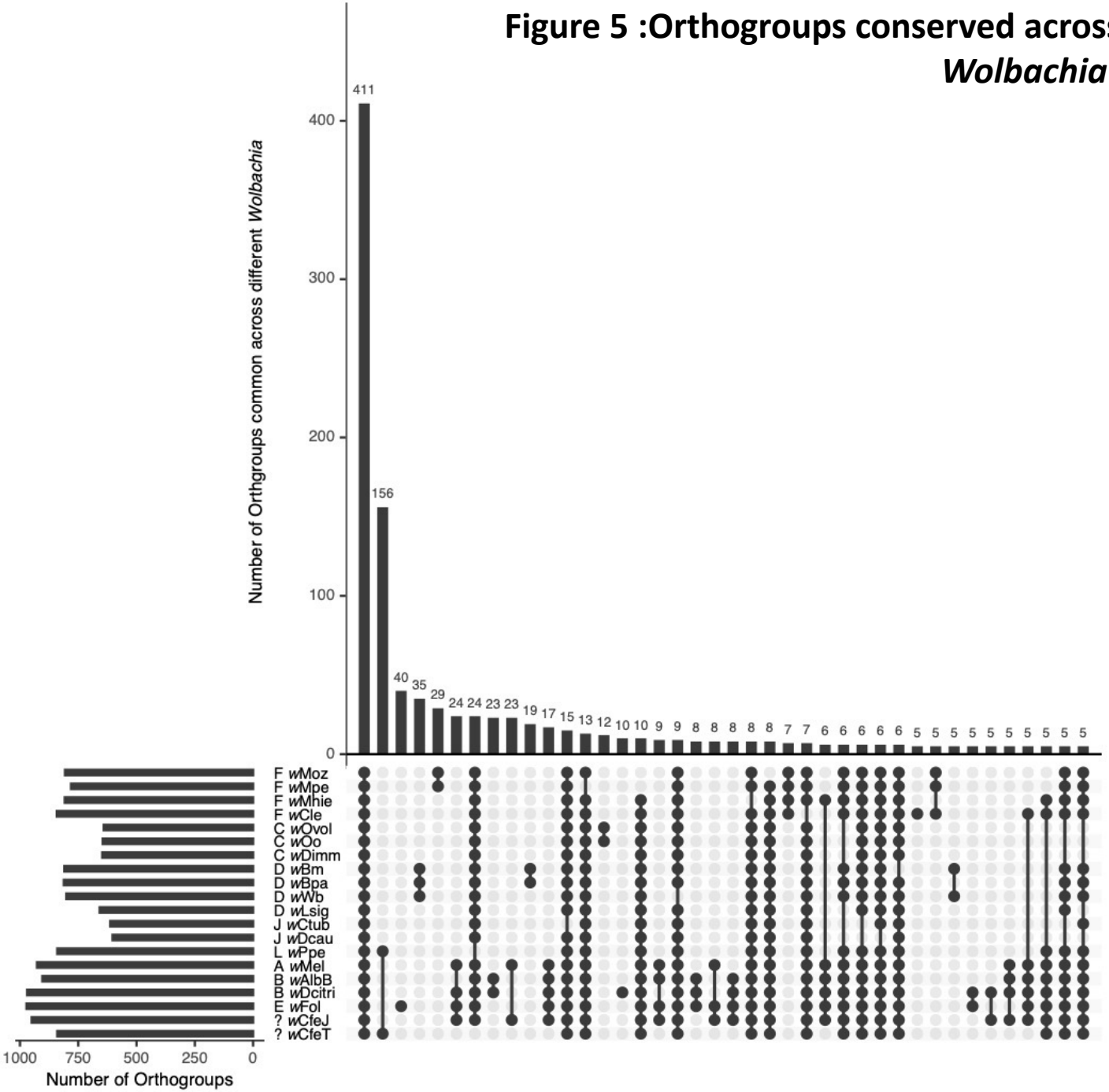
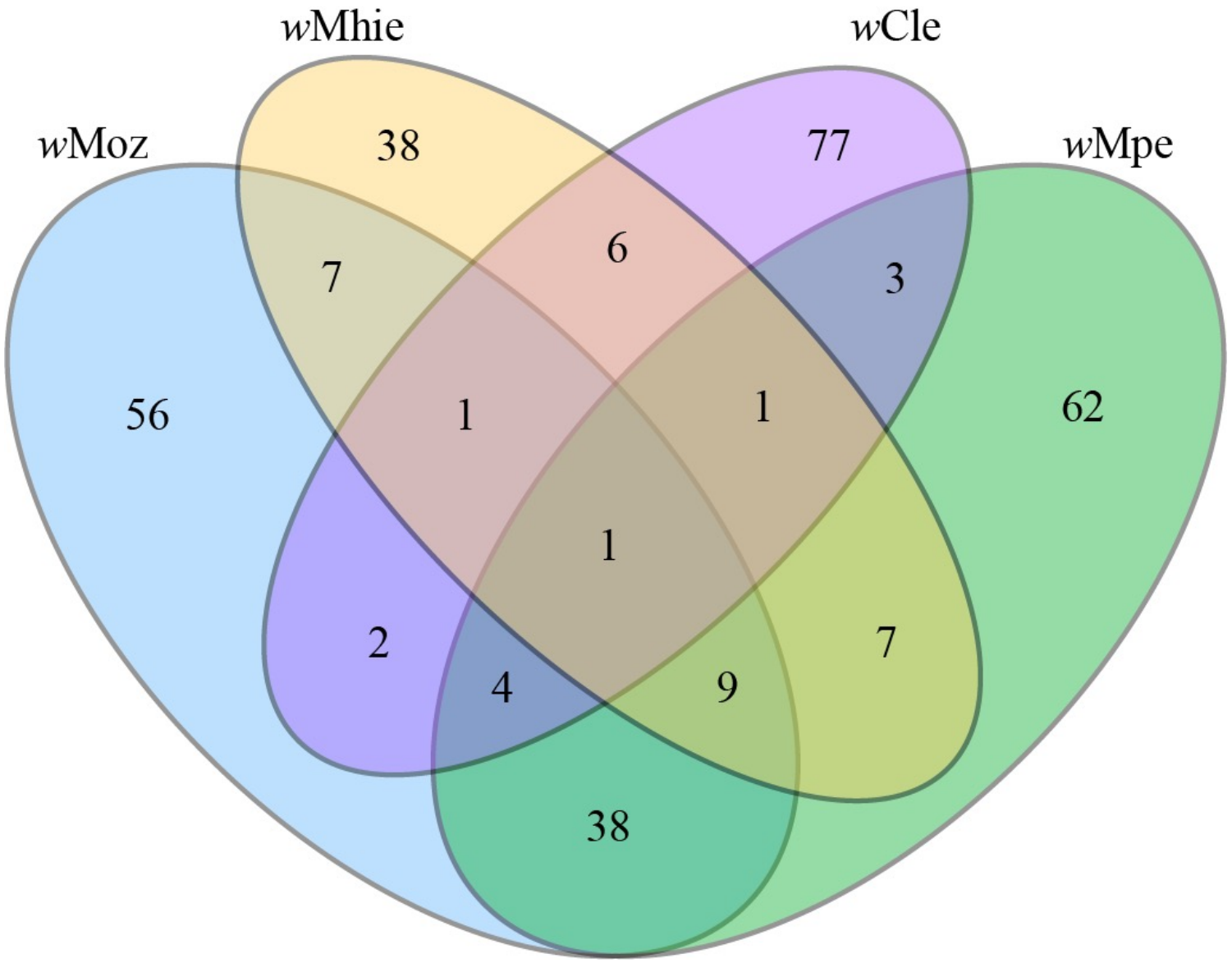


Figure 6. Venn diagram comparing pseudogenes in supergroup F *Wolbachia*



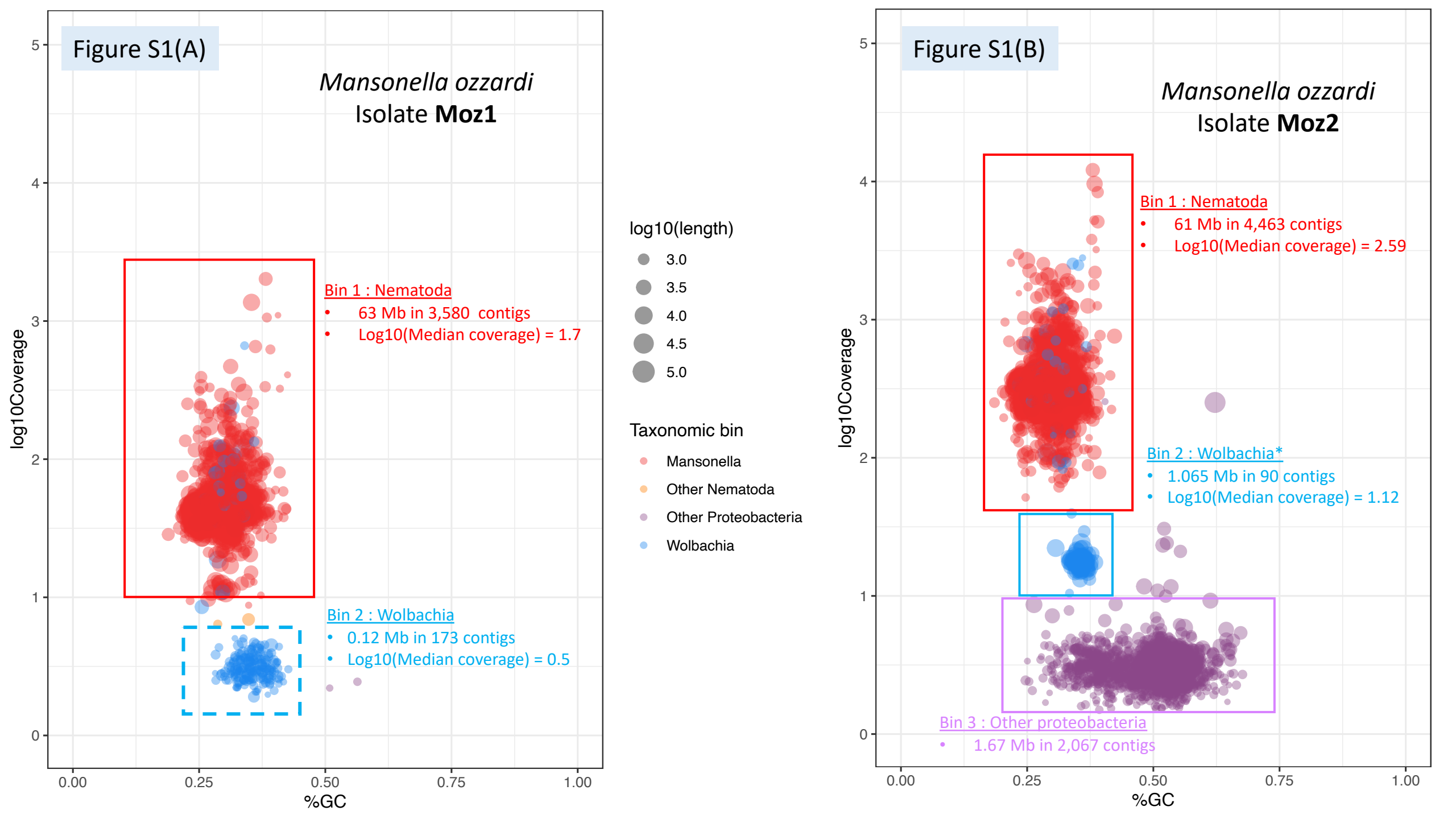


Figure S1(A)
Mansonella ozzardi
Isolate Moz1
Bin 1 : Nematoda
• 63 Mb in 3,580 contigs
• Log10(Median coverage) = 1.7
Bin 2 : Wolbachia
• 0.12 Mb in 173 contigs
• Log10(Median coverage) = 0.5
log10Coverage
%GC
0.00
0.25
0.50
0.75
1.00
0
1
2
3
4
5
log10(length)
3.0
3.5
4.0
4.5
5.0
Taxonomic bin
Mansonella
Other Nematoda
Other Proteobacteria
Wolbachia
Figure S1(B)
Mansonella ozzardi
Isolate Moz2
Bin 1 : Nematoda
• 61 Mb in 4,463 contigs
• Log10(Median coverage) = 2.59
Bin 2 : Wolbachia*
• 1.065 Mb in 90 contigs
• Log10(Median coverage) = 1.12
Bin 3 : Other proteobacteria
• 1.67 Mb in 2,067 contigs
log10Coverage
%GC

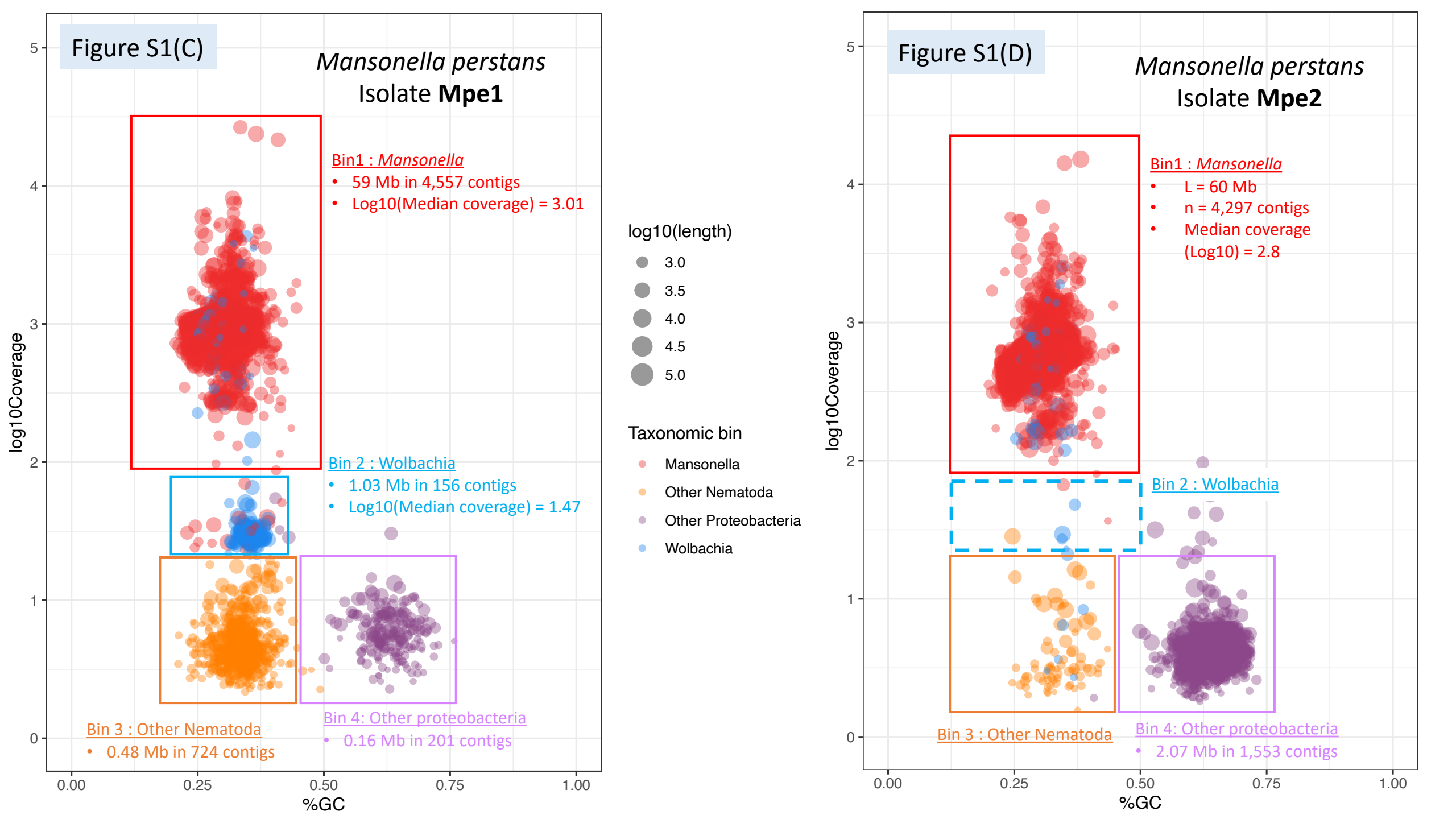
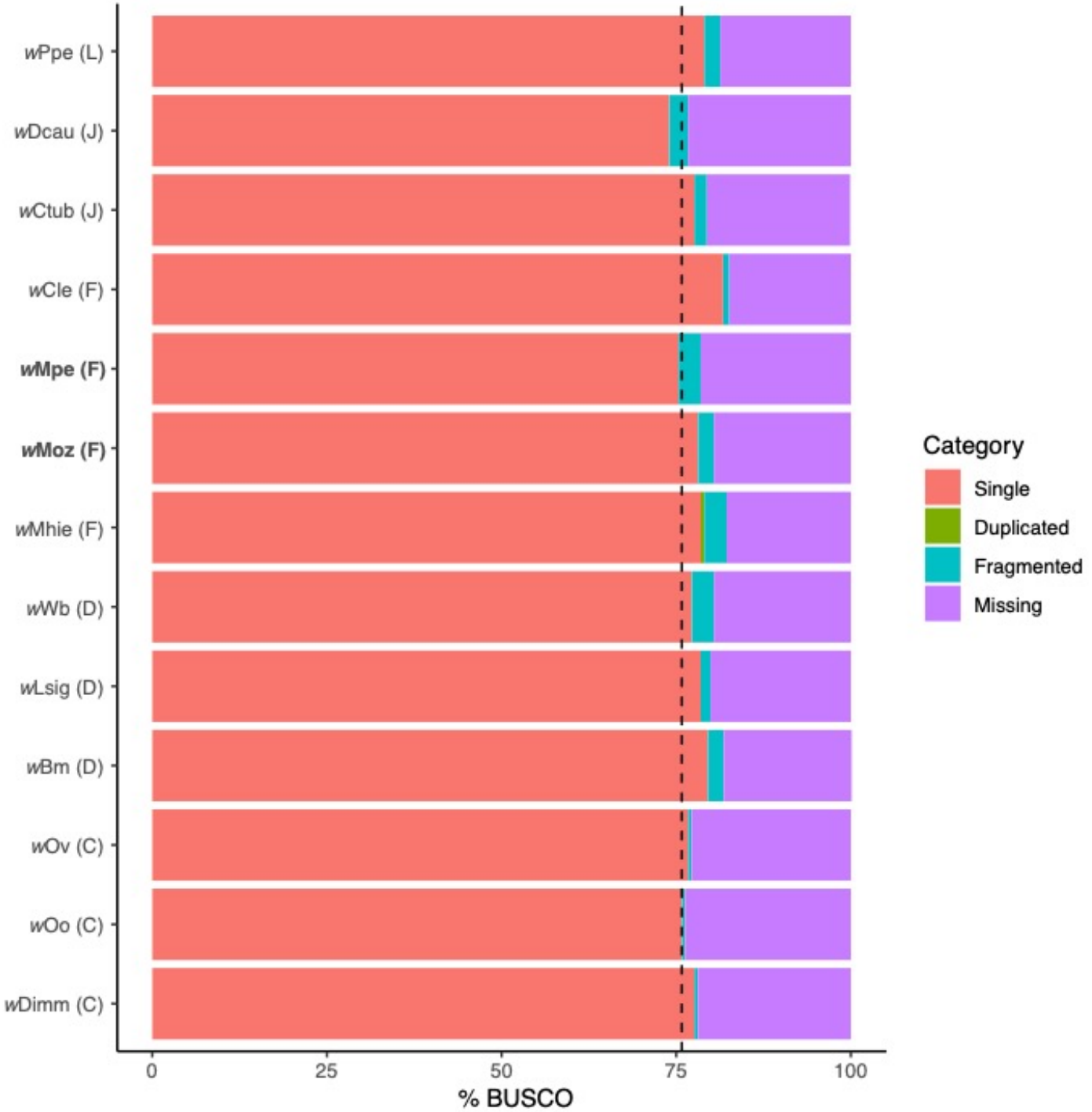


Figure S1(C)
Mansonella perstans
Isolate Mpe1
Bin1 : Mansonella
• 59 Mb in 4,557 contigs
• Log10(Median coverage) = 3.01
Bin 2 : Wolbachia
• 1.03 Mb in 156 contigs
• Log10(Median coverage) = 1.47
Bin 3 : Other Nematoda
• 0.48 Mb in 724 contigs
Bin 4: Other proteobacteria
• 0.16 Mb in 201 contigs
log10Coverage
%GC
0.00
0.25
0.50
0.75
1.00
0
1
2
3
4
5
log10(length)
3.0
3.5
4.0
4.5
5.0
Taxonomic bin
Mansonella
Other Nematoda
Other Proteobacteria
Wolbachia
Figure S1(D)
Mansonella perstans
Isolate Mpe2
Bin1 : Mansonella
• L = 60 Mb
• n = 4,297 contigs
• Median coverage (Log10) = 2.8
Bin 2 : Wolbachia
Bin 3 : Other Nematoda
Bin 4: Other proteobacteria
• 2.07 Mb in 1,553 contigs
log10Coverage
%GC
0.00
0.25
0.50
0.75
1.00
0
1
2
3
4
5

Figure S2. BUSCO scores of various *Wolbachia*

The BUSCO pipeline was used to measure the proportion of highly conserved, single copy orthologs (BUSCO groups). The set of reference BUSCO groups was set to the lineage "proteobacteria_odb10", which contains 219 BUSCOs derived from proteobacterial species. The *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.