




# Multiple Lineages of Nematode-*Wolbachia* Symbiosis in Supergroup F and Convergent Loss of Bacterioferritin in Filial *Wolbachia*

Amit Sinha <sup>1</sup>, Zhiru Li<sup>1</sup>, Catherine B. Poole<sup>1</sup>, Laurence Ettwiller <sup>1</sup>, Nathália F. Lima<sup>2</sup>, Marcelo U. Ferreira<sup>2</sup>, Fanny F. Fombad<sup>3</sup>, Samuel Wanji<sup>3</sup>, and Clotilde K.S. Carlow <sup>1,\*</sup>

<sup>1</sup>New England Biolabs, Ipswich, Massachusetts 01938

<sup>2</sup>Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, Brazil

<sup>3</sup>Department of Microbiology and Parasitology, University of Buea, Cameroon

\*Corresponding author: E-mail: carlow@neb.com.

Accepted: 24 April 2023

## Abstract

The intracellular endosymbiotic proteobacteria *Wolbachia* have evolved across the phyla nematoda and arthropoda. In *Wolbachia* phylogeny, supergroup F is the only clade known so far with members from both arthropod and filarial nematode hosts and therefore can provide unique insights into their evolution and biology. In this study, four new supergroup F *Wolbachia* genomes have been assembled using a metagenomic assembly and binning approach, wMoz and wMpe from the human filarial parasites *Mansonella ozzardi* and *M. perstans*, and wOcae and wMoviF from the blue mason bee *Osmia caerulescens* and the sheep ked *Melophagus ovinus*, respectively. A comprehensive phylogenomic analysis revealed two distinct lineages of filarial *Wolbachia* in supergroup F, indicating multiple horizontal transfer events between arthropod and nematode hosts. The analysis also reveals that the evolution of *Wolbachia*-filaria symbioses is accompanied by a convergent pseudogenization and loss of the bacterioferritin gene, a phenomenon found to be shared by all filarial *Wolbachia*, even those outside supergroup F. These observations indicate that differences in heme metabolism might be a key feature distinguishing filarial and arthropod *Wolbachia*. The new genomes provide a valuable resource for further studies on symbiosis, evolution, and the discovery of new antibiotics to treat mansonellosis.

**Key words:** *Wolbachia*, filaria, evolution, bacterioferritin, heme, phylogenomics.

## Introduction

*Wolbachia* are Gram-negative  $\alpha$ -proteobacteria of the order *Rickettsiales*, present as intracellular symbionts in many species of parasitic filarial nematodes and arthropods (Werren et al. 2008). Although the *Wolbachia* associations in arthropods range from reproductive parasites to nutritional mutualists (Werren et al. 2008; Zug and Hammerstein 2015), in filarial nematodes, *Wolbachia* is an obligate mutualist (Taylor et al. 2005) and essential for worm development, fecundity, and survival (Taylor et al. 2005; Pfarr and Hoerauf 2006). wMpe and wMoz are endosymbionts of *Mansonella perstans* and *M. 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 (Downes and Jacobsen 2010; Simonsen et al. 2011; Lima et al. 2016; Mediannikov and Ranque 2018; Ta-Tang et al. 2018; B elard and Gehringer 2021; Sandri et al. 2021). *M. perstans* is considered to be the most common filaria in West and Central Africa (Simonsen et al. 2011; Ta-Tang et al. 2018) and is also found in the Amazon rainforest and the Caribbean coast of South America (Tavares da Silva et al. 2017; Ta-Tang et al. 2018). *M. ozzardi* infections have been reported from many countries in South and Central America, including some Caribbean Islands (Lima et al. 2016; Raccurt 2018; Ta-Tang et al. 2018; Calvopina

  The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

## Significance

*Wolbachia* are bacterial endosymbionts of medically important parasitic filarial nematodes and arthropods. The evolutionary history and biological roles of *Wolbachia* in these different hosts are not well understood. The supergroup F in *Wolbachia* phylogeny harbors members from both filarial and arthropod hosts, providing an unparalleled opportunity for genomic comparisons to uncover distinguishing characteristics. This study provides new genomes from filarial and arthropod *Wolbachia* from this unique supergroup. Their phylogenomic analysis reveals multiple, independent transfers of *Wolbachia* between arthropod and filarial hosts. Remarkably, filaria–*Wolbachia* associations are associated with a convergent loss of the bacterioferritin gene, a key regulator of heme metabolism. Heme supplementation is considered a critical component of *Wolbachia*–filaria symbiosis. We demonstrate bacterioferritin loss is a novel feature exclusive to filarial *Wolbachia*.

et al. 2019). Coinfections of *M. perstans* and *M. ozzardi* are also known to exist in South America (Kozek et al. 1983; Crainey et al. 2020), complicating treatment as the two species respond differently to the commonly used antifilarial drugs (Ta-Tang et al. 2018). Antibiotics targeting *Wolbachia* have been successful in treating human filarial diseases such as onchocerciasis and lymphatic filariasis (Pfarr and Hoerauf 2006; Langworthy et al. 2000; Bazzocchi et al. 2008; Clare et al. 2019; Taylor et al. 2019; Hübner et al. 2019; Hong et al. 2019, 10). Doxycycline has been shown to clear *M. perstans* microfilariae (Keiser et al. 2008; Coulibaly et al. 2009; Ta-Tang et al. 2018) and may have some efficacy against *M. ozzardi*, as well as dual infections with *M. ozzardi* and *M. perstans* (Crainey et al. 2020); 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 with *Wolbachia* present in other human filarial parasites. The latest phylogenetic classification, based on multi-locus sequence typing (MLST), assigns *Wolbachia* to various monophyletic supergroups (Lefoulon et al. 2016; Lefoulon, Clark, Guerrero, et al. 2020a; Lefoulon, Clark, Borveto, et al. 2020b). Supergroups A, B, E, H, and S are comprised entirely of arthropod *Wolbachia*, whereas supergroups C, D, and J consist of filarial *Wolbachia* exclusively. Interestingly, the *Wolbachia* from *Mansonella* are classified under supergroup F, the only supergroup currently known to have *Wolbachia* members from filarial 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* (Hosokawa et al. 2010; Nikoh et al. 2014; Lefoulon et al. 2016; Lefoulon, Clark, Guerrero, et al. 2020a).

Genomic information provides a fundamental resource for these investigations, yet only two genomes from supergroup F are currently available in public databases, namely the genomes of *Wolbachia* wCle from the bedbug *Cimex lectularius* (Nikoh et al. 2014), and wMhie, from *Madathamugadia hiepei*, a filarial parasite of geckos

(Lefoulon, Clark, Guerrero, et al. 2020a). In the current study, the genomes of supergroup F *Wolbachia*, wMoz and wMpe from human filarial parasites *M. ozzardi* and *M. perstans* respectively, are assembled and analyzed. The genomes of arthropod-associated supergroup F *Wolbachia*, wOcae and wMoviF, from the mason bee, *Osmia caerulescens*, and sheep ked, *Melophagus ovinus*, are also reported. It is currently not known whether the *Wolbachia* of *M. ovinus* or *O. caerulescens* are facultative or obligate symbionts, or if they induce cytoplasmic incompatibility in these hosts. Genome-wide analyses of synteny, and annotations and comparative analysis of phage-derived gene content, cytoplasmic incompatibility genes and metabolic pathways of these genomes are described. In addition, comprehensive phylogenomic analyses of these four 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 distinct clades within supergroup F, each containing *Wolbachia* members from filarial as well as arthropod hosts, indicating multiple horizontal transfers between the two host phyla. Interestingly, the bacterioferritin gene was found to be pseudogenized in both clades of filarial *Wolbachia* in supergroup F, but not in their closest arthropod *Wolbachia*. The loss of bacterioferritin was discovered to be a feature common to all filarial *Wolbachia*.

## Results

### Genomes of wMoz and wMpe, Supergroup F *Wolbachia* From Human Filarial Parasites *M. ozzardi* and *M. perstans*

A metagenomic assembly and binning pipeline suitable for complex clinical samples were 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. S1a, Supplementary Material online). The BlobTools analysis of *M. ozzardi* isolate Moz2 from Venezuela (supplementary fig. S1b, Supplementary Material online) yielded a draft wMoz 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. S2, Supplementary Material online). The wMpe 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 wMoz and wMpe demonstrated high sequence similarity and colinearity 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 wCle, where high sequence synteny and colinearity were also observed (supplementary fig. S3, Supplementary Material online).

Gene prediction using the NCBI PGAP pipeline identified 888 protein coding genes in wMoz and 864 protein coding genes in wMpe. 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 wMoz and wMpe 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* wOv from the human filarial parasite *Onchocerca volvulus*, wOo from the bovine parasite *On. ochengi*, and wBm from the human filarial parasite *Brugia malayi* are 76.7%, 75.8%, and 79.5%, respectively (supplementary fig. S4, Supplementary Material online). Within supergroup F, the BUSCO score for filarial *Wolbachia* wMhie is 79%, and it is 81.7% for the arthropod *Wolbachia* wCle.

#### 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 scaffolds in the metagenomic assemblies (table 2). In the *M. ozzardi* sample Moz2, the median coverage of scaffolds classified as *Mansonella* was 360 $\times$ , whereas that of the *Wolbachia* scaffolds was 13 $\times$ . Assuming 1,000 nematode cells per

microfilaria (Basyoni and Rizk 2016), each parasite is estimated to harbor 37 *Wolbachia* cells. For the other *M. ozzardi* isolate Moz1, the median coverage for nematode scaffolds was 57 $\times$ , 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 1,032 $\times$  and 30 $\times$  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 a very small portion of the *Wolbachia* genome, 15 kb in 12 scaffolds (less than 2% of 1.073 Mb wMpe1 genome) with a median coverage of 7 $\times$  could be obtained during the metagenomic assembly, indicating a very low titer of *Wolbachia* in this isolate. The low titer was also confirmed using a complementary analysis, namely mapping the Mpe2 reads against the wMpe1 assembly. In this analysis, the median depth of coverage obtained was only 9 $\times$ , whereas the breadth of coverage (number of bases covered at 1 $\times$  depth of coverage) was observed to be 83 kb only, just 8% of the wMpe1 genome. Because 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* (Grobusch et al. 2003; Keiser et al. 2008; Coulibaly et al. 2009; Gehringer et al. 2014; Debrah et al. 2019; Sandri et al. 2021).

#### Genomes of wOcae and wMovIF, Supergroup F *Wolbachia* From the Arthropods *O. caerulescens* and *M. ovinus*

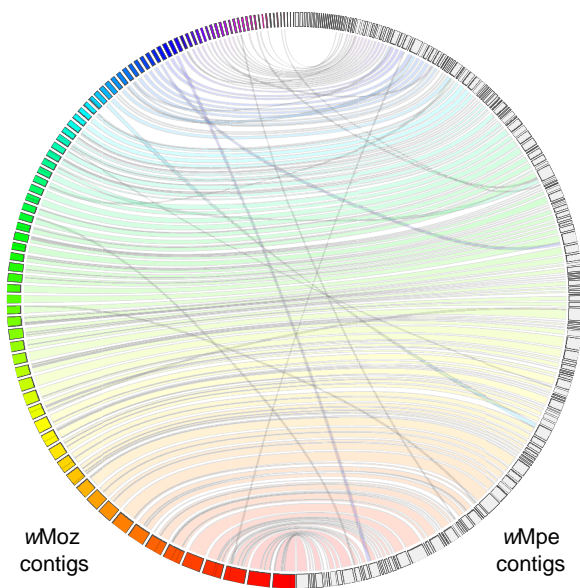
For the *Wolbachia* wOcae from the mason bee *O. caerulescens*, metagenomic assembly and binning were performed on Illumina sequencing reads available from a previous study (Gerth et al. 2014). A BlobTools analysis of the assembled scaffolds identified a distinct cluster of 186 *Wolbachia* scaffolds (supplementary fig. S5a, Supplementary Material online) with a total size of 1.2 Mb. A whole-genome alignment of these wOcae contigs against the wCle demonstrated high quality of this assembly (supplementary fig. S5b, Supplementary Material online). The NCBI-PGAP pipeline identified 899 protein coding genes, which had a BUSCO score of 79% (table 1). During the preparation of this manuscript, a metagenomic assembly WOLB0012 of *Wolbachia* from the same raw data was published (Scholz et al. 2020), but only the unannotated nucleotide sequence is available in NCBI database. A BlobPlot analysis of this assembly indicated the presence of arthropod contigs (supplementary fig. S6, Supplementary Material online). Therefore, instead of the

**Table 1**  
Characteristics of Genomes of wMoz and wMpe and Other *Wolbachia* From the Supergroup F

	wMoz	wMpe	wMoviF	wOcae	wMhie	wCle
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 reads	This study	This study	ERR969522	SRR122170	Lefoulon et al. 2020a, 2020b	Nikoh et al. 2014
Assembly accession and reference	GCF_020278625.1 This study	GCF_020278605.1 This study	Nováková et al. 2015 GCA_023661065.1 This study	Gerth et al. 2014 GCA_023661085.1 This study	GCF_013366855.1 Lefoulon et al. 2020a, 2020b	GCF_000829315.1 Nikoh et al. 2014

WOLB0012 assembly, the carefully curated and annotated version of the wOcae assembly described in this manuscript was used for all subsequent analyses.

The *Wolbachia* genome from the sheep ked *M. ovinus* was obtained from analysis of publicly available reads. A first round of metagenomic assembly and BlobTools analysis revealed a complex mixture of multiple *Wolbachia* as well as other proteobacterial endosymbionts (supplementary fig. S7, Supplementary Material online). 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* wCle, whereas others were more similar to the supergroup A *Wolbachia* wMel or supergroup B *Wolbachia* wAlbB. 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. S8a, Supplementary Material online). To identify the correct supergroup for each scaffold, its sequence similarity and query coverage in comparison with the wCle, wMel, or wAlbB genomes were analyzed. The median percentage identity as well as query-coverage scores was observed to be the highest against the wCle genome, followed by the wMel genomes, and was much lower when compared with the wAlbB genome (supplementary fig. S8b and c, Supplementary Material online). Therefore, to isolate the supergroup F scaffolds, alignments of each scaffold to wCle genome was compared with its alignments against the wMel genome. Among the 572 scaffolds, 40 scaffolds had a nucmer hit only to the wCle genome and were therefore included in the supergroup F bin, 214 scaffolds had a hit only to the wMel genome (supplementary fig. S8d, Supplementary Material online) and were included in the supergroup A bin, whereas 84 scaffolds that did not have a hit in either wCle or wMel 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 wCle and wMel genomes, an affiliation score, calculated as the difference between its percentage identity to the wCle genome and its percentage identity to the wMel genome, was analyzed. This score showed a clear bimodal distribution, with a range from -15 to 13 (supplementary fig. S8e, Supplementary Material online), 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 sequencing read-coverage



**Fig. 1.**—High-sequence similarity and synteny between the *de novo* assembled *wMoz* and *wMpe* genomes. The *Wolbachia* genomes assembled from two biologically and geographically distinct isolates display a high level of synteny and intergenome consistency. The genomes were aligned using minimap2 and visualized using the JupiterPlot software. The *wMoz* scaffolds are displayed as colored boxes in the left semi-circle, whereas the *wMpe* 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.

distribution similar to the 40 scaffolds that had exclusive hits to the *wCle* genome (supplementary fig. S8f, Supplementary Material online), confirming their classification into supergroup F. Finally, the classification of contigs as supergroup F was validated through an additional blobplot analysis using Illumina reads from a new metagenomic sequencing of a *M. ovinus* sample from China (Zhang et al. 2023). A combination of sequencing coverage from this new analysis and the affiliation scores described above achieved a clear classification (supplementary fig. S8g, Supplementary Material online). A whole-genome alignment of the contigs classified as *wMoviF* against the *wCle* demonstrated high quality of this assembly (supplementary fig. S8h, Supplementary Material online). The final supergroup F bin comprised of 196 scaffolds with a total size of 1.01 Mb was designated as the genome for *Wolbachia wMoviF*, 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. S4, Supplementary Material online). 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. Recently, a metagenomic assembly of *Wolbachia* (GenBank accession WOLB1015), based on the same raw reads, became available (Scholz et al. 2020). However, multiple analyses demonstrate that this assembly is a mixture of two different *Wolbachia*. First, the protein coding genes in WOLB1015 were predicted using PROKKA, and their BUSCO score was calculated. This analysis revealed duplications of 17% of single copy orthologs (SCOs) conserved across *Wolbachia*, an unusually high rate in contrast to 0% duplications that are observed in almost all other *Wolbachia* (supplementary fig. S4, Supplementary Material online). Second, a BlobPlot analysis of WOLB1015 showed the *Wolbachia* contigs spread over an unusually wide coverage range from  $1 \times$  to  $1000 \times$ , consistent with the presence of multiple *Wolbachia* in this assembly (supplementary fig. S9, Supplementary Material online). The WOLB1015 assembly had also failed to pass the QC criteria in the original study (Scholz et al. 2020). Therefore, instead of the mixed WOLB1015 assembly, the manually curated assembly of supergroup F *Wolbachia wMoviF* generated in this manuscript was used for all downstream analyses.

### Genome Sequence Comparisons between Various *Wolbachia*

Analysis of whole genome alignments of the four newly assembled genomes to the complete *wCle* genome showed that most of the regions absent from *wMoz*, *wMpe*, *wOcae*, or *wMhie* assemblies corresponded to regions containing insertion sequence (IS) element transposons in the *wCle* genome (fig. 2). This indicates that most of the protein coding regions have been recovered in these draft genome assemblies, and the missing regions largely correspond to the IS elements, which due to their repetitive nature present a technical challenge to assemble using only short-read data (Lefoulon et al. 2019; Sinha et al. 2019).

Comparisons of sequence similarity between all supergroup F *Wolbachia* genomes, and representatives from other supergroups including *Wolbachia* from filarial nematodes (Family Onchocercidae), a plant parasitic nematode *Pratylenchus penetrans* (Family Pratylenchidae), and arthropod hosts (supplementary table S1, Supplementary Material online) were performed using genome-wide average nucleotide identity (gANI, supplementary fig. S10a, Supplementary Material online) and digital DNA–DNA hybridization (dDDH) scores (supplementary fig. S10b–d, Supplementary Material online). The dDDH scores provide more sensitivity when comparing closely related species (Meier-Kolthoff et al. 2013). The *wMpe:wMoz* pair was found to have a dDDH score of 71.7% (supplementary fig. S10b, Supplementary Material online). For comparison, the dDDH scores for other closely related sympatric *Wolbachia* are: 96.1% for *wOo:wOv* pair from

**Table 2**Estimation of *Wolbachia* Levels per Microfilaria

Host species	Isolate name	Isolate Location	Median read coverage of <i>Mansonella</i> scaffolds (M)	Number and total size of assembled <i>Wolbachia</i> scaffolds	Number and total size of assembled <i>Wolbachia</i> scaffolds	<i>Wolbachia</i> genome assembled	Median read coverage of <i>Wolbachia</i> scaffolds (W)	Number of <i>Wolbachia</i> per microfilaria <sup>a</sup> (1000×W/M)
<i>M. ozzardi</i>	Moz1	Brazil	57	173 scaffolds	0.12 Mb	No	3	NA <sup>b</sup>
<i>M. ozzardi</i>	Moz2	Venezuela	360	93 scaffolds	1.073 Mb	wMoz	13	37
<i>M. perstans</i>	Mpe1	Cameroon	1,032	170 scaffolds	1.058 Mb	wMpe	30	30
<i>M. perstans</i>	Mpe2	Cameroon	633	5 scaffolds	8.781 kb	No	NA <sup>b</sup>	NA <sup>b</sup>

<sup>a</sup>Assuming 1000 cells per microfilaria.<sup>b</sup>Not calculated due to the very low number or coverage of *Wolbachia* scaffolds.

*Onchocerca* spp. and 93.3% for wBm:wBpa pair from *Brugia* spp. (supplementary fig. S10c and d, Supplementary Material online). These scores suggest wMpe and wMoz have diverged substantially from each other as compared to the wOo:wOv or wBm:wBpa pairs, most likely due to them having split from their common ancestor before the wBm:wBpa or wOo:wOv split.

#### Orthology Analysis, Gene Function, and Pathway Annotations

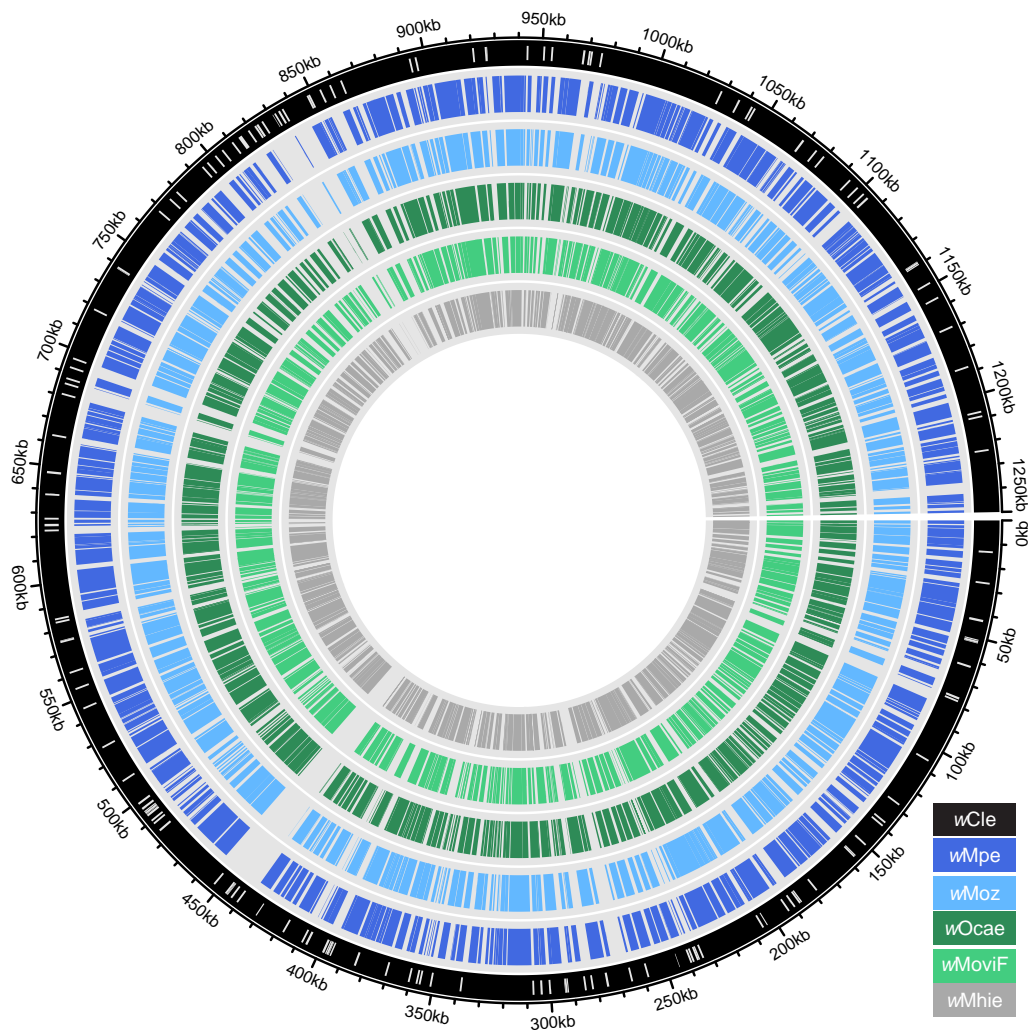
Protein-coding gene sequences from NCBI-PGAP annotations were obtained for all 123 *Wolbachia* genomes available in GenBank (supplementary table S1, Supplementary Material online). For orthology analysis, protein-coding genes from the four new genomes, in addition to 77 genomes selected to represent all supergroups, were used (see column H, supplementary table S1, Supplementary Material online). The OrthoFinder analysis of a total 82,406 proteins encoded by these 81 genomes identified 1,751 orthogroups, 99 of which represent SCOs conserved across all the analyzed *Wolbachia* (supplementary table S2, Supplementary Material online). For the four new supergroup F *Wolbachia*, at least 96.9% of all predicted proteins could be assigned to an orthogroup (supplementary table S2, Supplementary Material online).

Functional annotations of the proteins encoded by the wMoz, wMpe, wMoviF, and wOcae genomes were obtained using the eggNOG database (supplementary table S3, Supplementary Material online). Additionally, biosynthetic pathway annotations for the wCLe genome were obtained from the KEGG database, and corresponding orthologs were identified in the wMoz, wMpe, wMoviF, wOcae, and wMhie 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 wMoz and wMpe (supplementary table S4, Supplementary Material online). 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 (Raverdy et al. 2008), were also present in both wMoz and wMpe.

#### Biotin Biosynthesis Genes Are Absent From All Supergroup F *Wolbachia* Except wCLe

The bedbug *Wolbachia* wCLe genome harbors an operon encoding the enzymes involved in biosynthesis of biotin, which has been acquired via LGT from another *Rickettsia* (Nikoh et al. 2014). The biotin produced by wCLe supports a nutritional mutualism between the bedbug and its *Wolbachia* (Nikoh et al. 2014). 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 wMoz, wMpe, wMoviF, or wOcae proteomes. Additionally, when the sequencing reads from these four supergroup F *Wolbachia* were mapped against the wCLe 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. S11, Supplementary Material online). 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 six genes in the biotin pathway would have been detected. In contrast, the seven genes for biosynthesis of riboflavin, another vitamin pathway important for the host-symbiont relationship (Li and Carlow 2012; Moriyama et al. 2015; Ju et al. 2020) were present in wCLe as well as all other supergroup F *Wolbachia* (supplementary table S4, Supplementary Material online). 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.



**Fig. 2.**—Gaps in the assembled *Wolbachia* genomes correspond mostly to IS elements. Whole genome alignments of wMpe (blue ring), wMoz (light-blue ring), wOcae (green ring), wMoviF (light green ring), and wMhie (innermost gray ring) to the wCle chromosome (outermost black circle) are visualized as a circos plot. The white bars mark the locations of the IS elements in the wCle genome. The gaps in the alignments of various genomes to the wCle genome overlap extensively with IS elements in the wCle genome.

### 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 (Bordenstein and Bordenstein 2022). 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 (Bordenstein and Bordenstein 2022). Phage-derived genes in all supergroup F *Wolbachia* were identified via a comprehensive search for phage-derived orthogroups (fig. 3; [supplementary table S5, Supplementary Material](#) online) against a recently described database specific to *Wolbachia* phages (Bordenstein and Bordenstein 2022). In the wMoz and

wMpe 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* wMhie had 32 phage-associated orthogroups and shared 22 of these with wMoz and wMpe. Neither wMoz, wMpe, nor wMhie was found to possess any core components of phage structural modules, like the filarial *Wolbachia* from other supergroups. Among arthropod *Wolbachia* of supergroup F, wOcae had 29 phage structural genes, wMoviF had only three such genes, whereas wCle had only one, suggesting different dynamics of phage evolution in these *Wolbachia*.

Group II introns are common, phage-derived mobile elements (Bordenstein and Bordenstein 2022) that can be found in high copy numbers in many *Wolbachia* genomes: for example, 53 copies are found in the *wAlbB* genome (Sinha et al. 2019). Within the supergroup F genomes, the *wOcae* genome was found to encode only two copies of Group II introns. No such genes could be detected in the *wMoviF*, *wMoz*, or *wMpe* genome assemblies. One pseudogene was detected in both *wMoz* and *wMpe*. No intact gene or pseudogene for these elements could be found in the *wMhie* or *wCfeJ* genomes, as reported earlier (Lefoulon, Clark, Guerrero, et al. 2020a).

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 (Lindsey et al. 2018; Bordenstein and Bordenstein 2022). In supergroup F, no *cifA* or *cifB* genes could be detected in the genomes of arthropod *Wolbachia wCfeJ* or *wMoviF*. In *wOcae*, multiple pseudogenes or fragments of *cifA* and *cifB* genes missing either the 5-prime or 3-prime ends were identified (supplementary table S6, Supplementary Material online). Because 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 (Sinha et al. 2019). Thus, *wOcae* 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 wMoz*, *wMpe*, or *wMhie*.

### 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 all 123 annotated genomes available from GenBank (supplementary table S1, Supplementary Material online) as input for OrthoFinder, 46 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. S12, Supplementary Material online). 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 S1, Supplementary Material online). OrthoFinder analysis on these selected 81 genomes identified 99 SCOs (supplementary table S2, Supplementary Material online). Maximum likelihood trees with a partitioned model were built from the combined supermatrix of 62,160 nucleotides (fig. 4) and 20,470 amino acids

(supplementary fig. S13, Supplementary Material online). The four new genomes were robustly placed within supergroup F in this analysis. In this tree, the *Wolbachia wCfeJ* from the cat flea *C. felis*, which has not been assigned any supergroup yet, was found to be closest to supergroup F.

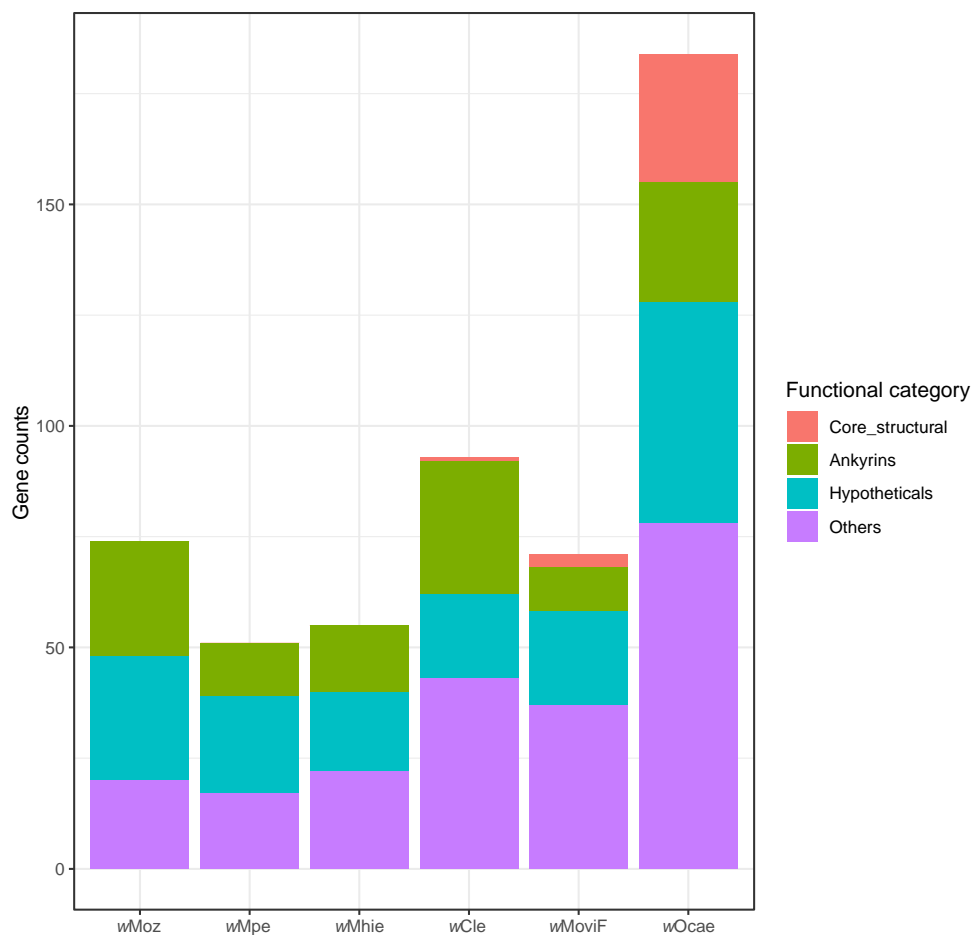
Remarkably, the filarial *Wolbachia wMoz* and *wMpe* were placed in a clade different from the clade harboring *wMhie*, the other filarial *Wolbachia* in this supergroup (fig. 5A). The filarial *Wolbachia wMoz* and *wMpe* were most closely related to the arthropod *Wolbachia wMoviF* and *wOcae*, whereas the filarial *Wolbachia wMhie* was more closely related to the arthropod *Wolbachia wCfeJ*. The topology of the *Wolbachia* phylogeny was also compared with the phylogeny of multiple *Wolbachia* hosts (fig. 5A), and limited congruence was found between the two phylogenies (fig. 5A). Particularly, within supergroup F, the clear separation of the filarial *Wolbachia* into two separate, well-supported clades, and the discordant phylogenetic topology of their corresponding hosts suggests at least two horizontal transfers of *Wolbachia* leading to these two clades.

To expand the analysis of supergroup F *Wolbachia* beyond those with whole genome sequences, a phylogenetic tree based on MLST loci (*ftsZ*, *fbpA*, *coxA*, *gatB*, *hcpA*) and *groEL*, *dnaA*, and 16S rRNA genes was constructed based on data available in GenBank (supplementary table S7, Supplementary Material online). This allowed including additional supergroup F *Wolbachia*, two from filarial hosts (*M. perforata* and *Cercopithifilaria japonica*) and four from different termite hosts. The corresponding sequences from *wCfeJ*, the *Wolbachia* found to be closest to supergroup F (fig. 5A), were used as an outgroup. A more distant *Wolbachia*, *wChemPL13* from supergroup T, which is a sister clade to supergroups F, C, J, S, and D was also included. This expanded tree (fig. 5B) also robustly supports the separation of the *Mansonella Wolbachia* clade from the *wMhie* clade. Interestingly, the only other known filarial *Wolbachia* in supergroup F, from the deer tick nematode *C. japonica*, is placed in a clade with *wMhie*, with the arthropod *Wolbachia wCfeJ* sharing a common ancestor with both these filarial *Wolbachia*. Together these observations demonstrate at least two distinct filarial *Wolbachia* lineages in supergroup F.

### Convergent Loss of Bacterioferritin across Filarial *Wolbachia*

Genome reduction and gene loss is a recurrent theme in the evolution of endosymbionts (Andersson and Kurland 1998), 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 have been



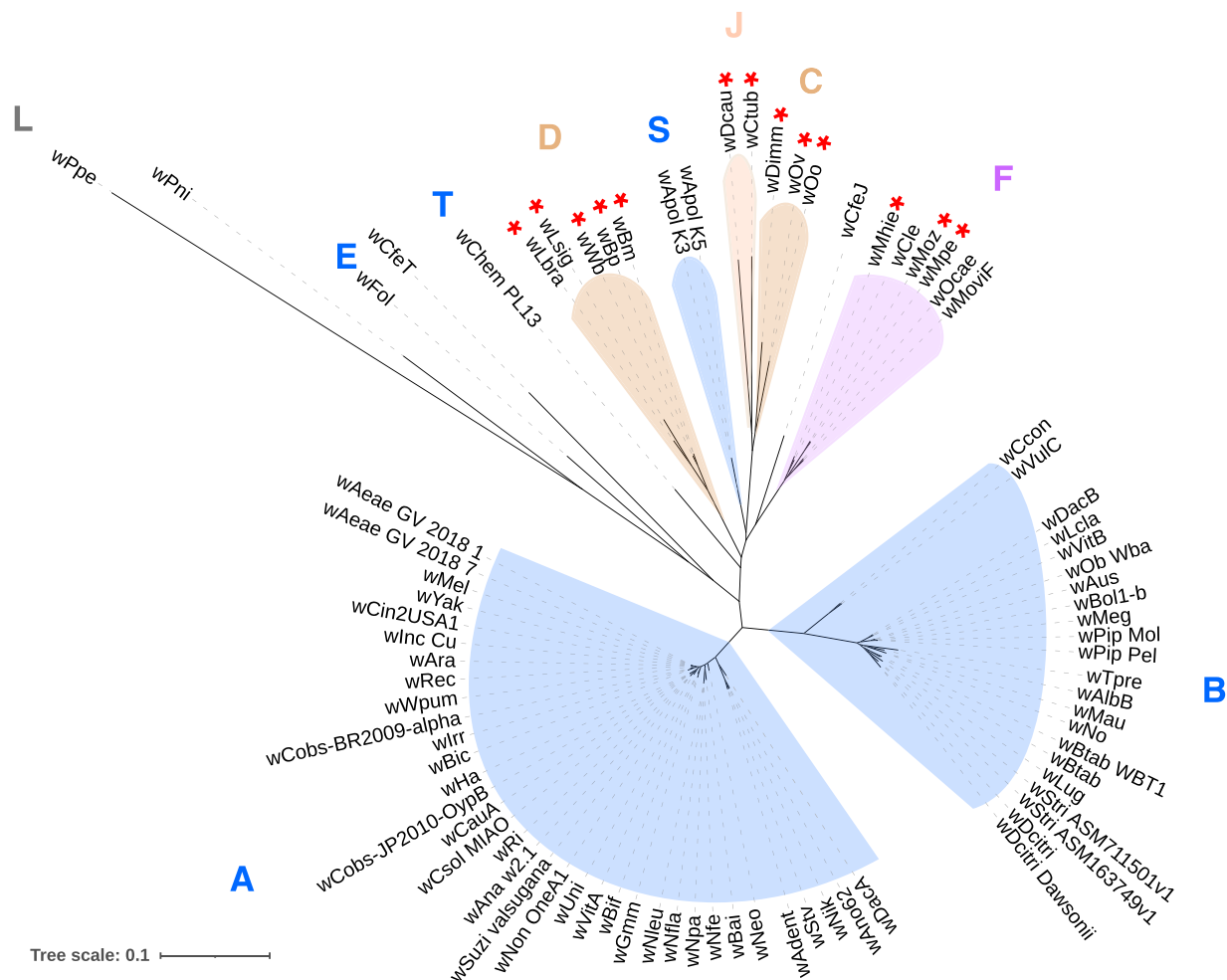


**Fig. 3.**—Numbers of phage-derived genes and their functional distributions in supergroup F *Wolbachia*. Phage-derived genes in each *Wolbachia* were categorized according to their predicted protein function. Phage-derived genes which had annotations other than “Hypothetical protein” but could not be classified as either “Core structural” or “Ankyrins” were included in the category “Others.” Only *wOcae* has intact phage modules, whereas *wMovIF* and *wCle* have only one and two core structural phage genes, respectively. Phage structural genes were absent from filarial *Wolbachia* *wMoz*, *wMpe*, and *wMhie*.

selected against during the evolution of filaria-*Wolbachia* symbiosis. The orthology relationships across 81 *Wolbachia* genomes (supplementary table S2, Supplementary Material online) were analyzed to look for orthogroups that were lost exclusively either in arthropod-associated or filaria-associated *Wolbachia*. Although no consistent orthogroup loss was found across arthropod *Wolbachia*, a striking loss of the bacterioferritin gene was observed in filarial *Wolbachia* across all supergroups C, D, J, and F (fig. 4). 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 *wMhie* clade (fig. 5A). These pseudogenes had missing canonical start codons and premature stop codons resulting in truncated proteins about 75% of the length of the intact protein (115 aa in *wMpe*, 125 aa in *wMoz*, 121 aa in *wMhie* as compared to 160 aa in *wCle*). The bacterioferritin proteins

functions as a complex of 24 subunits forming a spherical cage to trap heme molecules (Rivera 2017) and a truncated protein is likely to disrupt this complex structure. Within supergroup F, all filarial *Wolbachia*, namely *wMoz*, *wMpe*, and *wMhie*, contain a pseudogenized bacterioferritin, even though their most closely related arthropod *Wolbachia* have the corresponding gene intact. Because the *wMoz*:*wMpe* pair is placed in a separate clade from *wMhie* (fig. 5), our results suggest that the *bfr* gene has been pseudogenized independently in the two separate lineages.

Interestingly, pseudogene fragments of the *bfr* gene in various stages of degradation could be found across all sequenced filarial *Wolbachia* (supplementary fig. S15a, Supplementary Material online). 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



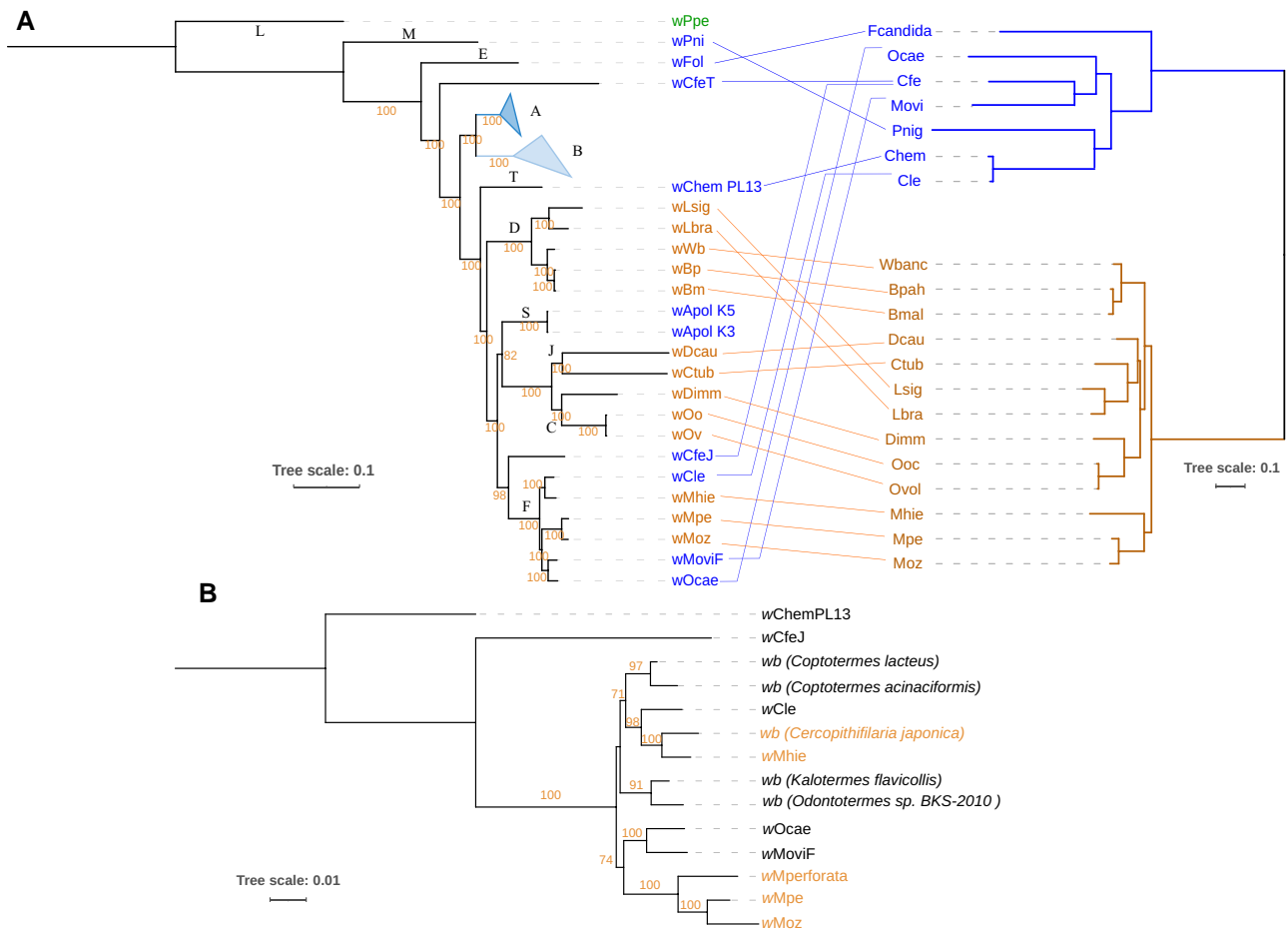
**FIG. 4.**—Global phylogenomic analysis of all *Wolbachia* genomes. Phylogenomic tree based on gene sequences of 91 SCOs conserved across 81 *Wolbachia* representing all supergroups. Supergroups comprised of *Wolbachia* exclusively from arthropod hosts are marked in blue, whereas the supergroups comprised 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.

region (supplementary fig. S15a, Supplementary Material online). This genomic region was observed to be syntenic across all analyzed genomes (supplementary fig. S15a, Supplementary Material online), confirming that the lack of the *bfr* gene is not due to assembly artifacts.

Although the bacterioferritin locus in the RefSeq annotations of *Wolbachia* wBpa from *B. pahangi* is annotated in NCBI 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 wBm, as well as other nucleotide insertions and deletions downstream (supplementary fig. S15b, Supplementary Material online), suggesting the pseudogenization of this gene in wBpa as well. In wBm, the *bfr* locus has a premature stop codon resulting in truncated protein with 138 amino acids instead of the complete

160 amino acids. The predicted peptide sequences corresponding to these pseudogenes showed multiple premature stop codons and frameshifts in both wBm and wBpa (supplementary fig. S15c, Supplementary Material online).

The dN/dS ratio, the ratio of nonsynonymous to synonymous substitutions for the *bfr* locus in wBpa, was found to be higher than the corresponding locus in wBm, (supplementary table S9, Supplementary Material online). Prior to dN/dS calculations, analysis of transitions and transversions with respect to genetic divergence plots and calculation of substitution saturation index (Xia et al. 2003) was performed to ensure that the bacterioferritin locus has not undergone substitution saturation across these *Wolbachia* (supplementary fig. S16, Supplementary Material online). Further, the dN/dS values for *bfr* locus in wBm and wBp were also compared against the distribution of genome-wide dN/dS ratios for all potential pseudogenes detected



**Fig. 5.**—Two independent lineages of *Wolbachia*-filial symbiosis within supergroup F. A) Mid-point rooted phylogenomic tree of *Wolbachia* based on gene sequences of 91 SCOs conserved across *Wolbachia* juxtaposed with a phylogenetic tree of various *Wolbachia* hosts based on 1,336 orthogroups. All filial species, their *Wolbachia* and the lines depicting host-*Wolbachia* relationships are in orange. The arthropod hosts, their *Wolbachia* and their connecting lines are in blue. The bootstrap values are shown along the respective branches. Supergroups A and B comprise of 35 and 22 *Wolbachia* genomes respectively and are represented together as blue triangles. *Wolbachia* supergroups are denoted near the root node of each supergroup. B) Phylogenetic tree of supergroup F *Wolbachia* based on a supermatrix of 8 genes (*ftsZ*, *fbpA*, *coxA*, *gatB*, *hcpA*, *groEL*, *dnaA*, 16S rRNA), total sequence length 10,374 nucleotides, also supports at least two independent lineages of filial *Wolbachia*. All *Wolbachia* from *Mansonella* hosts, including *M. perforata* are placed in a clade adjacent to the clade harboring arthropod *Wolbachia* *wOcae* and *wMoviF*, whereas the *Wolbachia* *wMhie* and *wCjaponica* from filial hosts *M. hiepei* and *C. japonica* are in a separate clade that is more closely related to the arthropod *Wolbachia* *wCle*. Filial *Wolbachia* are marked with orange font. Only the standard bootstrap values higher than 70 are shown.

in these genomes using Pseudofinder (Syberg-Olsen et al. 2022). This analysis showed the dN/dS values for the *bfr* locus in *wBm* and *wBpa* are substantially higher than the median values of genome-wide dN/dS (supplementary fig. S17, Supplementary Material online). Together, these analyses confirm that bacterioferritin is undergoing pseudogenization in all filial *Wolbachia*.

### 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 (Andersson and Kurland

1998). However, the observation of two independent *bfr* pseudogenization events within supergroup F *Wolbachia* suggests that additional factors, such as convergent evolution through selection against 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 S8, Supplementary Material online). These “pseudogene-orthogroups” were then compared with identify those unique or shared across different *Wolbachia* (fig. 6). 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 high number ( $n = 15$ ) of genes that were found to be pseudogenized in each of the three filarial *Wolbachia* *wMoz*, *wMpe*, and *wMhie*, whereas their closest sister arthropod *Wolbachia* had intact versions of the corresponding genes (fig. 6). The bacterioferritin pseudogene described above is one of these 15 genes. Because *wMoz* and *wMpe* are placed in a clade separate from *wMhie*, these 15 pseudogenes 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 *wMoz*:*wMpe* pair ( $n = 39$ ). The corresponding pseudogenes might have arisen either in the last common ancestor of *wMoz* and *wMpe* or through convergent evolution specific to *Wolbachia* within *Mansonella* or a combination of both factors.

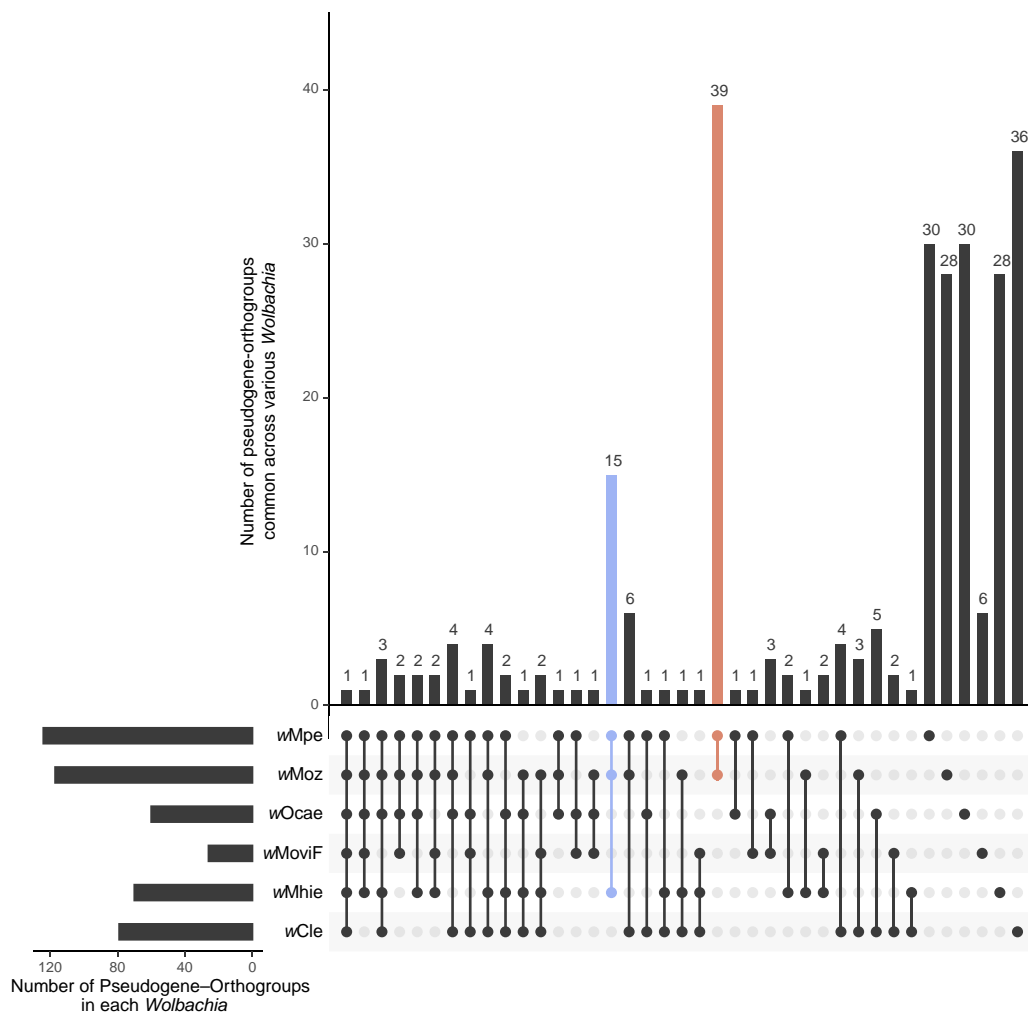
## Discussion

*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, four new *Wolbachia* genomes from supergroup F are described, expanding the number of genomes from two to six in this underrepresented supergroup. Two genomes are from *wMoz* and *wMpe*, endosymbionts of *Mansonella* spp., important filarial parasites of humans, whereas the other two genomes are from *wOcae* and *wMoviF*, endosymbionts of the mason bee *O. caerulea* and the sheep ked *M. ovinus* respectively.

Because *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 *wMoz* and *wMpe*, 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 *wOcae* and *wMoviF* genomes, the publicly available sequencing data were 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, were 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 (Scholz et al. 2020), and for highlighting sex-specific differences in *Wolbachia* levels in *D. immitis* (Kumar et al. 2013). Given the longstanding controversy around the detection of *Wolbachia* in *M. perstans* (Grobusch et al. 2003; Keiser et al. 2008; Coulibaly et al. 2009; Gehringer et al. 2014; Debrah et al. 2019; Sandri et al. 2021), determination of *Wolbachia* titers in *Mansonella* is particularly important and has implications for antibiotic use as a treatment (Gehringer et al. 2014). In the current study, the titer in the isolate Mpe1 was found to be 30 *Wolbachia* cells per microfilaria. In the isolate Mpe2, although the presence of *Wolbachia* sequences could be detected, their depth and breadth of coverage was too low for estimating the titer reliably. Importantly, this low *Wolbachia* coverage in Mpe2 is not simply due to an overall low sequencing depth, as the corresponding genome for the host *M. perstans* could be successfully assembled at a high coverage that was comparable to the isolate Mpe1. As both Mpe1 and Mpe2 samples were collected in Cameroon, these observations demonstrate that even within the same geographical region, *Wolbachia* levels can widely 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 Mpe1, and to levels reported from a laboratory strain of *B. malayi* (McGarry et al. 2004). The presence of *Wolbachia* in all *Mansonella* isolates described here, albeit at varying levels, and in another human parasite *Mansonella* spp. "DEUX" (Sandri et al. 2021), together with the successful clearance of *M. perstans* microfilariae with doxycycline (Coulibaly et al. 2009; Debrah et al. 2019), 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* because 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 (*wMoz*, *wMpe*, *wMoviF*, and *wOcae*) in supergroup F and identifies *wCfeJ* as the most closely related and robust outgroup to this supergroup. The phylogenomic analyses revealed two distinct clades of filarial *Wolbachia* within supergroup F. The *wMoz* and *wMpe* pair shared a common ancestor with the arthropod *Wolbachia* *wOcae*



**Fig. 6.**—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 *wMoz* and *wMpe* *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* *wMoz*, *wMpe*, and *wMhie* ( $n = 15$ , 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.

and *wMoviF*, whereas *wMhie* shared a different common ancestor with the bed bug *Wolbachia* *wCle*, indicating multiple events of horizontal transfer of *Wolbachia* between arthropod and filarial hosts. While this scenario has previously been proposed based on a few gene loci (Casiraghi et al. 2005; Lefoulon et al. 2012, 2016), the genome-wide phylogeny presented here, based on a supermatrix of 91 genes, with 2 new filarial and 2 new arthropod *Wolbachia* genomes, provides the most comprehensive and robust evidence in support of multiple horizontal transfers across the two host phyla. In addition to horizontal transfers across hosts, the widespread occurrence and transmission of *Wolbachia* infections across different animal phyla can also occur through cladogenic inheritance

or introgression. The relative contributions of these three transmission modes have been analyzed for *Wolbachia* in closely related host systems, such as in the *Nasonia* species complex (Raychoudhury et al. 2009) and different *Drosophila* species complex (Turelli et al. 2018; Cooper et al. 2019). Horizontal transfer of *Wolbachia* has been confirmed as the mechanism explaining very similar *Wolbachia* observed in the reproductively isolated *Drosophila* species *D. simulans* and *D. ananassae* (Turelli et al. 2018). The widely diverged arthropod and filarial hosts in supergroup F are also reproductively isolated species and phyla that harbor very similar *Wolbachia*. Therefore, horizontal transfer of *Wolbachia* between the two phyla is the most likely explanation for their observed phylogeny. This conclusion is

further supported by the incongruence between the phylogenies of supergroup F *Wolbachia* and their hosts, where *wMhie* and *wb C. japonica* are sister *Wolbachia* branching off the same node (fig. 5B), but their hosts are from different clades, ONC5 and ONC4, respectively (Lefoulon et al. 2016). Therefore, the *wMpe/wMoz* and the *wMhie/wb C. japonica* lineages are distinct and unlikely to be vertically inherited from a single filarial ancestor with an ancestral *Wolbachia*.

An interleaved pattern of clades with arthropod- or filarial-*Wolbachia* members is also evident beyond supergroup F, across the global phylogenomic analysis encompassing all available *Wolbachia*. The *Wolbachia* supergroups C and J with filarial hosts are a sister clade to the arthropod-only supergroup S. Deeper in the phylogeny, the filarial supergroup D shares a common ancestor with supergroups C, J, S, F, and *wCfeJ*. Furthermore, the supergroup T *Wolbachia wChemPL13* from a bed bug shares a common ancestor with the supergroup D, as well as the clade containing S, J, C, F, and *wCfeJ*. These phylogenetic patterns and the incongruence between the host-*Wolbachia* cophylogenies indicate multiple horizontal transfers between the arthropod and filarial hosts, but the direction of these transfers remains challenging to infer from the phylogenomic data alone.

The striking observation of a convergent pseudogenization of the bacterioferritin gene in filarial *Wolbachia* enables formulation of hypotheses regarding direction of *Wolbachia* transfer within supergroup F. At least two independent losses of the bacterioferritin gene were observed in this supergroup, one in the *wMoz-wMpe* clade and the other in the *wMhie* clade, whereas their sister *Wolbachia* from arthropod hosts had an intact *bfr* gene. This suggests that an intact *bfr* gene is likely the ancestral state, as a reversion from a pseudogene to an intact gene is very unlikely. This intact ancestral *bfr* gene could have been present in an ancestral *Wolbachia* that was associated with either a nematode or arthropod host. Because all the extant filarial *Wolbachia* with available genomes lack the *bfr* gene, the likelihood of a filarial nematode as an ancestral host is relatively low. This hypothesis is based on a limited number of filarial *Wolbachia* genomes, which harbor the *bfr* gene at various stages of pseudogenization (supplementary fig. S15a, Supplementary Material online). If future sampling of more diverse filarial *Wolbachia* identifies any such *Wolbachia* with intact bacterioferritin, the current hypothesis will no longer be applicable.

The loss of bacterioferritin has important implications for the obligate relationship between *Wolbachia* and filarial hosts. Bacterioferritin is known to be 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 (Carrondo 2003). In arthropods, the presence of *Wolbachia* provides a fitness

benefit by increasing fecundity during conditions of iron deficiency (Brownlie et al. 2009), while during iron excess, *Wolbachia* protect their insect hosts from iron induced toxicity by upregulating bacterioferritin (Kremer et al. 2009). 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 (Foster et al. 2005; Gill et al. 2014). 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. Under this model, the various ancestral filarial *Wolbachia* might have had an intact bacterioferritin gene to begin with, but in the lineages that lost this gene due to genetic drift, the freeing up of heme for the filarial host would have provided a strong enough fitness advantage to drive the *bfr*-lacking *Wolbachia* to fixation in filarial hosts and initiating an obligate relationship. Although genome reduction and gene loss, usually via genetic drift, is a recurrent theme in the evolution of endosymbionts (Andersson and Kurland 1998), the loss of bacterioferritin gene across multiple filarial *Wolbachia* is a unique example of convergent gene loss accompanying the adaptation to a new ecological niche. Further research into the role of bacterioferritin and heme metabolism can better elucidate their role in different host-*Wolbachia* systems. In summary, this study provides four newly assembled supergroup F *Wolbachia* genomes. A global phylogenomic analysis including all sequenced *Wolbachia* genomes together with these new genomes has provided strong support for multiple lineages of filarial *Wolbachia* in supergroup F, as well as uncovered an intriguing loss of bacterioferritin from all filarial *Wolbachia*. Further studies of these four new genomes are expected to shed more light on the evolution and nature of symbiosis, as well as the identification of new treatments for mansonellosis.

## 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 (De Almeida Basano et al. 2018; Lima et al. 2018, 39).

### 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 (Lima et al. 2018). 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 1,989 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>) and then stored at  $-20^{\circ}\text{C}$ . 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 (Poole et al. 2019). Mpe1 DNA was extracted using a Genomic DNA Tissue MicroPrep kit (Zymo Research, USA) as described previously (Poole et al. 2019). 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 enrichment process is based on selective binding and removal of CpG-methylated DNA islands present in human genome but absent from nematode and *Wolbachia* genomes. The relative amounts of DNA from nematode and *Wolbachia* therefore remains unaffected after the enrichment step. The preparation of Illumina libraries from Mpe1 and Moz1 samples has been described as part of a previous

study (Poole et al. 2019), 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 and 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 (Nurk et al. 2017). Input reads were mapped back to the assembled scaffolds using bowtie2 v.2.4.1 (Langmead and Salzberg 2012). Binning of metagenomic scaffolds into metagenome assembled genomes of *Mansonella* and *Wolbachia* was performed using the BlobTools v1.1.1 software (Laetsch and Blaxter 2017) 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 *wCle* 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 *wMoz* and *wMpe* 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 wOcae* from the mason bee *O. caerulescens* was assembled from Illumina reads available under NCBI SRA accession SRR1221705, originally described in a previous study of arthropod *Wolbachia* (Gerth et al. 2014). 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 wMoviF* from the sheep ked *M. ovinus* was assembled from raw Illumina reads corresponding to the NCBI SRA accession number ERR969522, which originated from a study of sheep ked endosymbionts (Nováková et al. 2015) that were also used recently for *Wolbachia* genome mining (Scholz et al. 2020). 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 a contaminating genome from 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 *wCle* (supergroup F), *wMel* (supergroup A), and *wAlbB* (supergroup B) were performed using nucmer v4.0.0beta2 (Marçais et al. 2018, 4) with the parameter “mincluster” set to 100.

### Genome Annotation and Comparative Analysis

For both *wMpe* and *wMoz* genomes, as well as the *wOcae* and *wMoviF* genomes, protein-coding genes, rRNA, tRNA,

ncRNA genes, and pseudogenes were identified using the NCBI prokaryotic genome annotation pipeline (Tatusova et al. 2016). The completeness of the protein-coding content of all four genomes was assessed using the BUSCO pipeline v5.1.3 using the “proteobacteria\_odb10” reference dataset (Simão et al. 2015). 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 (Li 2018). The parameters used in Jupiter plots are maxGap = 100,000; minBundleSize = 1,000; m\_ref\_contig\_cutoff = 500; and 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 *wMoz*, *wMpe*, *wOcae*, and *wMoviF*, as well as *wMhie*, against *wCle* 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 (Krzywinski et al. 2009) drawn using the R package circlize v0.4.10 (Gu et al. 2014). For global sequence comparisons across multiple *Wolbachia* genomes, the average nucleotide identity (ANI) scores were calculated using the OrthoANIu tool v1.2 (Yoon et al. 2017). 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*, dDDH scores (Meier-Kolthoff et al. 2013) 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 (Emms and Kelly 2019).

IS elements were identified via the ISSaga web server (Varani et al. 2011) 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 (Huerta-Cepas et al. 2017) web server (<http://eggnogetdb.embl.de/#/app/emapper>). The analysis of different metabolic pathways was based on the annotations of *wCle* and *wBm* genomes available in the KEGG database (Kanehisa et al. 2017).

### 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 S1,



Supplementary Material online) 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, for example multiple isolates of the same *Wolbachia* (supplementary table S1, Supplementary Material online). The third set included only the genomes from supergroup F *Wolbachia*, and *wCfeJ* *Wolbachia* from the cat flea *Ctenocephalides felis* as an outgroup.

For each input set, OrthoFinder was used to identify SCOs conserved across all genomes within the set. Protein sequences corresponding to each orthogroup were aligned using mafft v7.149b (Nakamura et al. 2018). The multiple sequence alignments were trimmed using Gblocks v0.91b (Talavera and Castresana 2007). 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 SCOs 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 (Chernomor et al. 2016) as inputs to iqtree v2.1.2 (Nguyen et al. 2015). Automatic selection of substitution models was performed using ModelFinder (Kalyaanamoorthy et al. 2017) implementation in iqTree (iqTree command line option “-m MFP”). Evaluation of branch supports was performed using the iqTree implementations of 1) the SH-like approximate likelihood ratio test (Guindon et al. 2010) and 2) ultrafast bootstrap (Hoang et al. 2018) with 1,000 replicates each (iqTree command-line options “-B 1000 -alrt 1000 -abayes -lbp 1000”). The consensus trees were visualized using the iTOL webserver (Letunic and Bork 2021) and further annotated in Adobe Illustrator.

For a phylogenetic analysis of all supergroup F *Wolbachia* based on a combined supermatrix of multiple genes, the sequences for their MLST loci *ftsZ*, *fbpA*, *coxA*, *gatB*, *hcpA*, and other loci (*groEL*, *dnaA*, *ftsZ* and 16S rRNA) were obtained from their genome assemblies and GenBank. The corresponding sequences from *wCfeJ* was used as outgroup. Only the *Wolbachia* with sequences available for at least three of these eight loci were included in this analysis. The multiple sequence alignment of the combined supermatrix of sequences was performed using MUSCLE (Edgar 2004). Phylogenetic tree with automatic model selection, standard bootstrap calculations, and single branch

tests was generated using iqtree (command-line options -b 1000 -alrt 1000 -abayes).

### Assembly and Annotation of *Wolbachia* Host Genomes and Transcriptomes for Phylogenomic Analysis

The protein coding genes of different *Wolbachia* hosts were obtained as follows. The genomes and proteomes for filarial nematodes *M. perstans*, *M. ozzardi* were recently described (Sinha et al. 2023). Curated proteomes from publicly available data for the filarial hosts *B. malayi*, *B. pahangi*, *Wuchereria bancrofti*, *Dirofilaria immitis*, *On. ochengi*, *On. volvulus*, *Litomosoides sigmodontis*, *L. brasiliensis*, *Dipetalonema caudispina*, and *Cruorifilaria tubero cauda* were the same as recently (Sinha et al. 2023). The genome sequence of *O. caerulescens* was obtained from the metagenomic assembly and blobplot analysis for *wOcae* described above. Annotated genes for *C. lectularius* and *Folsomia candida* were obtained from NCBI assemblies GCF\_000648675.2 and GCF\_002217175.1 respectively. The *Pentalonia nigronervosa* genome was obtained from NCBI accession GCA\_014851325.1. Modeling and masking of genomic repeats in *O. caerulescens* and *P. nigronervosa* genomes was performed using RepeatModeler version 2.0.3 (Flynn et al. 2020) and Dfam database (Storer et al. 2021). These repeats were soft-masked in the respective genomes using RepeatMasker 4.1.2-p1 (<http://repeatmasker.org/>). Gene predictions were carried out on WebAUGUSTUS online server () with the soft-masked genome as input. The soft-masked genomes obtained were used as input to gene prediction using online AUGUSTUS web server (Hoff and Stanke 2013). The transcriptomes for *M. ovinus* and *C. hemipterus* were assembled de novo from SRA datasets SRR17267914 (Zhang et al. 2023) and SRR19217017 respectively. The reads were corrected using rcorrector (Song and Florea 2015) and assembled using Trinity-v2.8.5 (Haas et al. 2013). The longest super-transcripts were selected using transdecoder v5.5.0 (<https://github.com/TransDecoder/TransDecoder>), and arthropod sequences were selected using kraken version 2.0.9-beta (Wood and Salzberg 2014) and kraken-tools script extract\_kraken\_reads.py (Lu et al. 2022). Conserved orthologs and species tree for these nematode and arthropod hosts were obtained from OrthoFinder (Emms and Kelly 2019).

### 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 (Conway et al. 2017). Pseudogenes derived from transposons and IS elements were excluded from this analysis.

The dN/dS values for *wBpa-bfr* and *wMhie-bfr* loci were calculated from pairwise alignments with intact genes from closely related *Wolbachia*, using PAL2NAL utility (Suyama et al. 2006). The substitution saturation index of both these loci was evaluated using Xia’s test (Xia et al. 2003) using DAMBE (Xia 2018). The proportion of invariant sites at *bfr* locus for this analysis was obtained via PhyML (Guindon et al. 2010) analysis of 13 intact bacterioferritin genes from all supergroups. A global analysis of dN/dS values for all genes and pseudogenes for *wBpa* and *wBma* against the *wChemPL13* reference genome were obtained using Pseudofinder (Syberg-Olsen et al. 2022).

### 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 (Bordenstein and Bordenstein 2022). The sequences of *Wolbachia* phage genes and pseudogenes were obtained based on the Phage-WO maps recently described (Bordenstein and Bordenstein 2022). 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 (Martinez et al. 2021) 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*.

### Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

### 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., and S.W. collected the clinical samples. A.S., Z.L., C.B.P., and L.E. performed the sequencing experiments and bioinformatics analysis. A.S., Z.L., C.B.P., L.E., and C.K.S.C. analyzed the data. A.S., Z. L., C.B. P., and 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.

### Data Availability

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 PRJNA666671 and PRJNA666672, respectively. The assembled genome and PGAP annotations for the *Wolbachia* *wMpe* and *wMoz* are available from the NCBI GenBank database with accession numbers JACZH U000000000 and JADAKK000000000, respectively. The annotated genomes for *wOcae* and *wMovif* are available as GenBank accessions JAHDJT000000000 and JAHR BA000000000, respectively, deposited under BioProject accessions PRJNA244005 and PRJNA739984, respectively.

### Literature Cited

- Andersson SGE, Kurland CG. 1998. Reductive evolution of resident genomes. *Trends Microbiol.* 6:263–268.
- Basyoni MMA, Rizk EMA. 2016. Nematodes ultrastructure: complex systems and processes. *J Parasit Dis.* 40:1130–1140.
- Bazzocchi C, et al. 2008. Combined ivermectin and doxycycline treatment has microfilaricidal and adulticidal activity against *Dirofilaria immitis* in experimentally infected dogs. *Int J Parasitol.* 38: 1401–1410.

- Bélaré S, Gehringer C. 2021. 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.
- Bordenstein Sarah R, Bordenstein Seth R. 2022. Widespread phages of endosymbionts: phage WO genomics and the proposed taxonomic classification of symbiiviridae. *PLoS Genet.* 18:e1010227.
- Brownlie JC, et al. 2009. Evidence for metabolic provisioning by a common invertebrate endosymbiont, *Wolbachia pipientis*, during periods of nutritional stress. *PLoS Pathog.* 5:e1000368.
- Calvopina M, Chiluisa-Guacho C, Toapanta A, Fonseca D, Villacres I. 2019. High prevalence of *Mansonella ozzardi* infection in the Amazon region, Ecuador. *Emerg Infect Dis.* 25:2081–2083.
- Carrondo MA. 2003. Ferritins, iron uptake and storage from the bacterioferritin viewpoint. *EMBO J.* 22:1959–1968.
- Casiraghi M, et al. 2005. Phylogeny of *Wolbachia pipientis* based on *gltA*, *groEL* and *ftsZ* gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the *Wolbachia* tree. *Microbiology* 151: 4015–4022.
- Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol.* 65:997–1008.
- Clare RH, et al. 2019. Industrial scale high-throughput screening delivers multiple fast acting macrofilaricides. *Nat Commun.* 10:11.
- Conway JR, Lex A, Gehlenborg N, Hancock J. 2017. Upsetr: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33:2938–2940.
- Cooper BS, Vanderpool D, Conner WR, Matute DR, Turelli M. 2019. *Wolbachia* acquisition by *Drosophila yakuba*-clade hosts and transfer of incompatibility loci between distantly related *Wolbachia*. *Genetics* 212:1399–1419.
- Coulibaly YI, et al. 2009. A randomized trial of doxycycline for *Mansonella perstans* infection. *N Engl J Med.* 361:1448–1458.
- Crainey JL, et al. 2020. Deep sequencing reveals occult mansonellosis coinfections in residents from the Brazilian Amazon village of São Gabriel da Cachoeira. *Clin Infect Dis.* 71:1990–1993.
- De Almeida Basano S, et al. 2018. 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.
- Debrah LB, et al. 2019. The efficacy of doxycycline treatment on *Mansonella perstans* infection: an open-label, randomized trial in Ghana. *Am J Trop Med Hyg.* 101:84–92.
- Downes B, Jacobsen K. 2010. A systematic review of the epidemiology of mansonellosis. *Afr J Infect Dis.* 4:1–14.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32: 1792–1797.
- Emms DM, Kelly S. 2019. Orthofinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238.
- Flynn JM, et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A.* 117: 9451–9457.
- Foster J, et al. 2005. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol.* 3:e121.
- Gehringer C, et al. 2014. Molecular evidence of *Wolbachia* endosymbiosis in *Mansonella perstans* in Gabon, Central Africa. *J Infect Dis.* 210:1633–1638.
- Gerth M, Gansauge M-T, Weigert A, Bleidorn C. 2014. Phylogenomic analyses uncover origin and spread of the *Wolbachia* pandemic. *Nat Commun.* 5:5117.
- Gill AC, Darby AC, Makepeace BL. 2014. Iron necessity: the secret of *Wolbachia*'s success? *PLoS Negl Trop Dis.* 8:e3224.
- Grobusch MP, Kombila M, Autenrieth I, Mehlhorn H, Kreamsner PG. 2003. No evidence of *Wolbachia* endosymbiosis with *Loa loa* and *Mansonella perstans*. *Parasitol Res.* 90:405–408.
- Gu Z, Gu L, Eils R, Schlesner M, Brors B. 2014. Circlize implements and enhances circular visualization in R. *Bioinformatics* 30:2811–2812.
- Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59:307–321.
- Haas BJ, et al. 2013. De novo transcript sequence reconstruction from RNA-Seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 8:1494–1512.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 35:518–522.
- Hoff KJ, Stanke M. 2013. WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Res.* 41:W123–W128.
- Hong WD, et al. 2019. AWZ1066S, a highly specific anti-*Wolbachia* drug candidate for a short-course treatment of filariasis. *Proc Natl Acad Sci U S A.* 116:1414–1419.
- Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T. 2010. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proc Natl Acad Sci U S A.* 107:769–774.
- Hübner MP, et al. 2019. In vivo kinetics of *Wolbachia* depletion by ABBV-4083 in *L. sigmodontis* adult worms and microfilariae. *PLoS Negl Trop Dis.* 13:e0007636.
- Huerta-Cepas J, et al. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol.* 34:2115–2122.
- Ju J-F, et al. 2020. *Wolbachia* supplement biotin and riboflavin to enhance reproduction in planthoppers. *ISME J.* 14:676–687.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 14:587–589.
- Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45:D353–D361.
- Keiser PB, et al. 2008. Molecular identification of *Wolbachia* from the filarial nematode *Mansonella perstans*. *Mol Biochem Parasitol.* 160:123–128.
- Kozek WJ, Palma G, Henao A, García H, Hoyos M. 1983. 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.
- Kremer N, et al. 2009. *Wolbachia* interferes with ferritin expression and iron metabolism in insects. *PLoS Pathog.* 5:e1000630.
- Krzywinski M, et al. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res.* 19:1639–1645.
- Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M. 2013. Blobology: exploring raw genome data for contaminants, symbionts, and parasites using taxon-annotated GC-coverage plots. *Bioinforma Comput Biol.* 4:237.
- Laetsch DR, Blaxter ML. 2017. Blobtools: interrogation of genome assemblies. *F1000Res.* 6:1287.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 9:357–359.
- Langworthy NG, et al. 2000. 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.
- Lefoulon E, et al. 2012. A new type F *Wolbachia* from splendofilariinae (onchocercidae) supports the recent emergence of this supergroup. *Int J Parasitol.* 42:1025–1036.

- Lefoulon E, et al. 2016. Breakdown of coevolution between symbiotic bacteria *Wolbachia* and their filarial hosts. *PeerJ* 4:e1840.
- Lefoulon E, et al. 2019. Large enriched fragment targeted sequencing (LEFT-SEQ) applied to capture of *Wolbachia* genomes. *Sci Rep.* 9: 1–10.
- Lefoulon E, Clark T, Borveto F, et al. 2020b. Pseudoscorpion *Wolbachia* symbionts: diversity and evidence for a new supergroup. *BMC Microbiol.* 20:188.
- Lefoulon E, Clark T, Guerrero R, et al. 2020a. Diminutive, degraded but dissimilar: *wolbachia* genomes from filarial nematodes do not conform to a single paradigm. *Microb Genom.* 6:e000487.
- Letunic I, Bork P. 2021. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49:W293–W296.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.
- Li Z, Carlow CKS. 2012. Characterization of transcription factors that regulate the type IV secretion system and riboflavin biosynthesis in *Wolbachia* of *Brugia malayi*. *PLoS One.* 7:e51597.
- Lima NF, Aybar CAV, Juri MJD, Ferreira MU. 2016. *Mansonella ozzardi*: a neglected new world filarial nematode. *Pathog Glob Health.* 110:97–107.
- Lima NF, Gonçalves-Lopes RM, Kruize YCM, Yazdanbakhsh M, Ferreira MU. 2018. CD39 and immune regulation in a chronic helminth infection: the puzzling case of *Mansonella ozzardi*. *PLoS Negl. Trop. Dis.* 12:e0006327.
- Lindsey ARI, et al. 2018. Evolutionary genetics of cytoplasmic incompatibility genes *cifA* and *cifB* in prophage WO of *Wolbachia*. *Genome Biol. Evol.* 10:434–451.
- Lu J, et al. 2022. Metagenome analysis using the kraken software suite. *Nat Protoc.* 17:2815–2839.
- Marçais G, et al. 2018. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol.* 14:e1005944.
- Martinez J, Klasson L, Welch JJ, Jiggins FM. 2021. Life and death of selfish genes: comparative genomics reveals the dynamic evolution of cytoplasmic incompatibility. *Mol Biol Evol.* 38:2–15.
- McGarry HF, Egerton GL, Taylor MJ. 2004. Population dynamics of *Wolbachia* bacterial endosymbionts in *Brugia malayi*. *Mol Biochem Parasitol.* 135:57–67.
- Mediannikov O, Ranque S. 2018. Mansonellosis, the most neglected human filariasis. *New Microbes New Infect.* 26:S19–S22.
- Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics.* 14:60.
- Moriyama M, Nikoh N, Hosokawa T, Fukatsu T. 2015. Riboflavin provisioning underlies *Wolbachia*'s fitness contribution to its insect host. *mBio* 6:e01732-15.
- Nakamura T, Yamada KD, Tomii K, Katoh K. 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* 34:2490–2492.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating Maximum-likelihood phylogenies. *Mol Biol Evol.* 32:268–274.
- Nikoh N, et al. 2014. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. *Proc Natl Acad Sci U S A.* 111:10257–10262.
- Nováková E, Husník F, Šochová E, Hypša V. 2015. *Arsenophonus* and *Sodalis* symbionts in louse flies: an analogy to the *Wigglesworthia* and *Sodalis* system in tsetse flies. *Appl Environ Microbiol.* 81: 6189–6199.
- Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27: 824–834.
- Pfarr KM, Hoerauf AM. 2006. 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.
- Poole CB, et al. 2019. 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.
- Raccurt CP. 2018. *Mansonella ozzardi* and its vectors in the new world: an update with emphasis on the current situation in Haiti. *J Helminthol.* 92:655–661.
- Raverdy S, Foster JM, Roopenian E, Carlow CKS. 2008. The *Wolbachia* endosymbiont of *Brugia malayi* has an active pyruvate phosphate dikinase. *Mol Biochem Parasitol.* 160:163–166.
- Raychoudhury R, Baldo L, Oliveira DCSG, Werren JH. 2009. Modes of acquisition of *Wolbachia*: horizontal transfer, hybrid introgression, and codivergence in the *nasonia* Species Complex. *Evolution* 63: 165–183.
- Rivera M. 2017. Bacterioferritin: structure, dynamics, and protein-protein interactions at play in iron storage and mobilization. *Acc Chem Res.* 50:331–340.
- Sandri TL, et al. 2021. Molecular epidemiology of *Mansonella* species in Gabon. *J Infect Dis.* 223:287–296.
- Scholz M, et al. 2020. Large scale genome reconstructions illuminate *Wolbachia* evolution. *Nat Commun.* 11:5235.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
- Simonsen PE, Onapa AW, Asio SM. 2011. *Mansonella perstans* filariasis in Africa. *Acta Trop.* 120:S109–S120.
- Sinha A, et al. 2023. Genomes of the human filarial parasites *Mansonella perstans* and *Mansonella ozzardi*. *Front Tropical Dis.* 4. doi: 10.3389/ftid.2023.1139343.
- Sinha A, Li Z, Sun L, Carlow CKS. 2019. Complete genome sequence of the *Wolbachia* wAlbB endosymbiont of *Aedes albopictus*. *Genome Biol Evol.* 11:706–720.
- Song L, Florea L. 2015. Rcorrector: efficient and accurate error correction for illumina RNA-Seq reads. *GigaScience* 4:48.
- Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. 2021. The dfam community resource of transposable element families, sequence models, and genome annotations. *Mob DNA.* 12:2.
- Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34:W609–W612.
- Syberg-Olsen MJ, Garber AI, Keeling PJ, McCutcheon JP, Husník F. 2022. Pseudofinder: detection of pseudogenes in prokaryotic genomes. *Molecular Biol Evol.* 39:msac153.
- Ta-Tang T-H, Crainey JL, Post RJ, Luz SL, Rubio JM. 2018. Mansonellosis: current perspectives. *Res Rep Trop Med.* 9:9–24.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 56:564–577.
- Tatusova T, et al. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44:6614–6624.
- Tavares da Silva LB, et al. 2017. Molecular verification of new world *Mansonella perstans* parasitemias. *Emerg Infect Dis.* 23:545–547.
- Taylor MJ, et al. 2019. Preclinical development of an oral anti-*Wolbachia* macrolide drug for the treatment of lymphatic filariasis and onchocerciasis. *Sci Transl Med.* 11:eaau2086.
- Taylor MJ, Bandi C, Hoerauf A. 2005. *Wolbachia* bacterial endosymbionts of filarial Nematodes. In: Baker JR, Muller R, Rollinson D, editors. *Advances in parasitology*. Vol. 60. Cambridge, MA: Academic Press. p. 245–284. doi: 10.1016/S0065-308X(05)60004-8.
- Turelli M, et al. 2018. Rapid global spread of wRi-like *Wolbachia* across multiple *Drosophila*. *Curr Biol.* 28:963–971.e8.
- Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. 2011. 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.
- Werren JH, Baldo L, Clark ME. 2008. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol.* 6:741–751.
- Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15:R46.
- Xia X. 2018. DAMBE7: new and improved tools for data analysis in molecular biology and evolution. *Mol Biol Evol.* 35:1550–1552.
- Xia X, Xie Z, Salemi M, Chen L, Wang Y. 2003. An index of substitution saturation and its application. *Mol Phylogenet Evol.* 26:1–7.
- Yoon S-H, Ha S, Lim J, Kwon S, Chun J. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek.* 110:1281–1286.
- Zhang Q, et al. 2023. The genome of sheep ked (*Melophagus ovinus*) reveals potential mechanisms underlying reproduction and narrower ecological niches. *BMC Genomics.* 24:1–11.
- Zug R, Hammerstein P. 2015. Bad guys turned nice? A critical assessment of *Wolbachia* mutualisms in arthropod hosts. *Biol Rev.* 90:89–111.

**Associate editor:** Daniel Sloan