

# *Wolbachia* endosymbionts in two *Anopheles* species indicates independent acquisitions and lack of prophage elements

Shannon Quek<sup>1</sup>, Louise Cerdeira<sup>2</sup>, Claire L. Jeffries<sup>3</sup>, Sean Tomlinson<sup>2</sup>, Thomas Walker<sup>3,\*</sup>, Grant L. Hughes<sup>1,2,\*</sup> and Eva Heinz<sup>2,4,\*</sup>

## Abstract

*Wolbachia* is a genus of obligate bacterial endosymbionts that infect a diverse range of arthropod species as well as filarial nematodes, with its single described species, *Wolbachia pipientis*, divided into several 'supergroups' based on multilocus sequence typing. *Wolbachia* strains in mosquitoes have been shown to inhibit the transmission of human pathogens, including *Plasmodium* malaria parasites and arboviruses. Despite their large host range, *Wolbachia* strains within the major malaria vectors of the *Anopheles gambiae* and *Anopheles funestus* complexes appear at low density, established solely on PCR-based methods. Questions have been raised as to whether this represents a true endosymbiotic relationship. However, recent definitive evidence for two distinct, high-density strains of supergroup B *Wolbachia* within *Anopheles demeilloni* and *Anopheles moucheti* has opened exciting possibilities to explore naturally occurring *Wolbachia* endosymbionts in *Anopheles* for biocontrol strategies to block *Plasmodium* transmission. Here, we utilize genomic analyses to demonstrate that both *Wolbachia* strains have retained all key metabolic and transport pathways despite their smaller genome size, with this reduction potentially attributable to degenerated prophage regions. Even with this reduction, we confirmed the presence of cytoplasmic incompatibility (CI) factor genes within both strains, with wAnD maintaining intact copies of these genes while the *cifB* gene was interrupted in wAnM, so functional analysis is required to determine whether wAnM can induce CI. Additionally, phylogenetic analysis indicates that these *Wolbachia* strains may have been introduced into these two *Anopheles* species via horizontal transmission events, rather than by ancestral acquisition and subsequent loss events in the *Anopheles gambiae* species complex. These are the first *Wolbachia* genomes, to our knowledge, that enable us to study the relationship between natural strain *Plasmodium* malaria parasites and their anopheline hosts.

## DATA SUMMARY

Sequence data generated and used for this analysis are available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) (BioProject number accession no. PRJNA642000). The two assembled *Wolbachia* genomes are available with genome accession numbers GCA\_018491735.2 and GCA\_018491625.2. Additional *Wolbachia* genomes used for comparative analysis are described in the supplementary material. Additional supplementary data files used to generate several figures can be found on FigShare (<https://doi.org/10.6084/m9.figshare.19576432>) [1].

Received 18 November 2021; Accepted 03 March 2022; Published 21 April 2022

**Author affiliations:** <sup>1</sup>Department of Tropical Disease Biology, Centre for Neglected Tropical Diseases, Liverpool School of Tropical Medicine, Liverpool, UK; <sup>2</sup>Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool, UK; <sup>3</sup>Department of Disease Control, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK; <sup>4</sup>Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK.

**\*Correspondence:** Thomas Walker, [thomas.walker@lshmt.ac.uk](mailto:thomas.walker@lshmt.ac.uk); Grant L. Hughes, [grant.hughes@lstmed.ac.uk](mailto:grant.hughes@lstmed.ac.uk); Eva Heinz, [eva.heinz@lstmed.ac.uk](mailto:eva.heinz@lstmed.ac.uk)

**Keywords:** *Anopheles*; genomics; prophage; symbiosis; *Wolbachia*.

**Abbreviations:** CI, cytoplasmic incompatibility; EAM, eukaryotic association module; ENA, European Nucleotide Archive; IS, insertion sequence; KAAS, KEGG Automatic Annotation Server; NCBI, National Center for Biotechnology Information; SNV, single nucleotide variant; T4SS, type IV secretion system.

The NCBI SRA BioProject accession number for the sequence data generated and used for this analysis is PRJNA642000. The GenBank/EMBL/DDBJ accession numbers for the two assembled *Wolbachia* genomes are GCA\_018491735.2 and GCA\_018491625.2.

**Data statement:** All supporting data, code and protocols have been provided within the article or through supplementary data files. Seven supplementary figures and one supplementary table are available with the online version of this article.

000805 © 2022 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

## Impact Statement

*Wolbachia* naturally infects a wide range of arthropod species, including insect vectors of human pathogens, where they may play a role in inhibiting their replication. These bacteria have been commonly found within *Aedes albopictus* and *Culex pipiens* mosquitoes, but have been noticeably absent in the mosquito genus *Anopheles*, which includes all species responsible for malaria transmission. Recent PCR-based methods have suggested the potential for natural *Wolbachia* strains within the *Anopheles gambiae* species complex, which includes major malaria vector species such as *Anopheles gambiae* s.s., *Anopheles coluzzii* and *Anopheles arabiensis*. We recently reported the presence of stable *Wolbachia* strains naturally occurring within two different *Anopheles* species (*Anopheles demeilloni* and *Anopheles moucheti*). In this study, we update and perform comparative genomic analysis of these two *Wolbachia* genomes against each other and published *Wolbachia* strains. Updated assemblies indicate smaller genome sizes compared to other *Wolbachia* of insects, despite their metabolic pathway repertoire being comparable to other strains. Interestingly, prophage fragments were identified within only one of the two strains. The findings of this study will be of significant interest to researchers investigating *Wolbachia* as a potential malaria biocontrol strategy, giving greater insight into the evolution and diversity of this obligate intracellular endosymbiont.

## INTRODUCTION

*Wolbachia* has a wide host range, including insects [2] where various estimates have predicted 52–60% of all arthropod species are naturally infected [3, 4]. Attempts to characterize the within-species diversity has resulted in the designation of *Wolbachia* ‘supergroups’ A through to T [5, 6], with several exceptions [7], via multilocus sequence typing of five single-copy conserved genes [8]. The relationship between *Wolbachia* and their hosts can range from obligate mutualism, where the endosymbiont is essential for host survival and reproduction [9, 10], to reproductive parasitism, where it manipulates the reproduction of its host to spread through the population. Currently, the best-studied phenotype (which also affects mosquito hosts) is cytoplasmic incompatibility (CI), which causes infected males to produce unviable offspring unless they mate with an infected female, while infected females have viable offspring regardless of the males’ infection status; thus, conferring a fitness advantage to *Wolbachia*-infected females.

Genetic studies have previously identified a pair of CI factor genes, *cifA* and *cifB* [11, 12], that have been correlated to this phenotype. These genes have often been found to co-occur as a single operon within prophage eukaryotic association modules (EAMs), and are believed to spread via horizontal transmission between *Wolbachia* strains due to their localization within prophage regions [13]. Despite being part of the same operon, these genes have been observed to be differentially regulated, with *cifA* having higher expression relative to *cifB* [14]. When these CI genes were first identified, they were placed into three distinct phylogenetic groups. While all three were recognized to maintain protein domains with predicted nuclease activity, the catalytic residues for these nuclease domains were predicted to be absent in one of the three groups [11], which instead contained an additional protein domain with ubiquitin-like specific protease activity [12, 15, 16]. This was later characterized as the type I group [15]. Additionally, recent research has identified genes encoding similar features in other members of the order *Rickettsiales*, often found associated with mobile genetic elements, such as plasmids [15, 17]. As a result, a recent study has identified up to five phylogenetic types, with one of these types being identifiable in other *Rickettsia* as well as *Wolbachia* [15].

Utilization of the CI phenotype has been explored as the basis for potential mosquito control strategies to reduce human disease transmission. The bacterium is capable of inducing CI in both natural [6, 18, 19] and artificially infected lines [20–22], and possible methods to utilize them for mosquito population control include release of males infected with *Wolbachia* [23], or potentially via release of genetically modified mosquitoes that carry the CI genes, but not *Wolbachia* [24, 25]. In addition to inducing the CI phenotype, *Wolbachia* has been shown to interfere with pathogen replication directly, both in those that cause disease in the insect, as well as human pathogens that utilize the insect as a vector [6, 26–28]. This has been observed to be most effective with artificial infections of *Wolbachia* in non-native host mosquitoes [20–22]. Recent trials have shown that *Wolbachia* can be used to great effect in preventing the spread of dengue virus [29, 30], while laboratory trials have indicated their potential to block *Plasmodium* replication in artificially infected *Anopheles* mosquitoes [31–33]. While there is little evidence to date of stable *Wolbachia* infections [31], a stable infection within *Anopheles stephensi* is possible [32, 33]. Infection of these mosquitoes with *Wolbachia* was associated with significantly reduced hatch rates however [32, 34], possibly affecting the viability of CI as a control tool in this system.

Despite *Wolbachia*’s presence in a wide variety of insects, natural high-density strains within the mosquito genus *Anopheles* have not been conclusively proven [35, 36] until recently [37]. Previous efforts to detect this bacterium required highly sensitive PCR techniques [38–42] that amplify a select handful of *Wolbachia* genes. Unfortunately, this alone cannot confirm the presence of live bacteria or stable *Wolbachia* strains within insects. Furthermore, phylogenetic placement of these amplified *Wolbachia* sequences within *Anopheles gambiae* shows multiple strains distributed across supergroups A and B, with some strains not assigned to any supergroup [36].

We recently demonstrated high-density *Wolbachia* strains in two *Anopheles* species, *Anopheles demeilloni* and *Anopheles moucheti* [37, 43], which we observed in wild populations collected over a large geographical range in temporally distinct populations. Importantly, we further visualized these bacteria in the germline, as well as sequenced near-complete *Wolbachia* genomes from both host species [37]. Here, we present reassembled, circularized genomes for both strains, as well as in-depth comparative analyses of these two *Wolbachia* strains against each other, and in the broader context of *Wolbachia* supergroups A through to F, with specific focus on supergroup B. We show that, in terms of both size and predicted protein-encoding genes, both assembled genomes are at the low end of the range of *Wolbachia* strains found within insects whilst containing reduced or, in the case of *wAnM*, no prophage WO regions. Despite this, both *Wolbachia* genomes maintained complete pathways that are expected for the genus, such as complete haem and nucleotide biosynthetic pathways and type IV secretion systems (T4SSs). Additionally, we reconstructed the phylogenetic history using whole-genome-sequence data, which indicates that these strains may originate from independent acquisitions via horizontal transfer events, and not from an ancestral infection that has since been lost in other *Anopheles* mosquitoes.

## METHODS

### Sequence data collection and genome quality assessment

Both genome assemblies of *wAnD* and *wAnM* were manually curated (i.e. gaps, indels and synteny) using the approach described by Tsai and collaborators in 2010 [44], Mummer/Nucmer software tool v4.0.0 [45], Mauve v2.4.0 [46] and Tablet v1.21.02.08 [47]. To complement the genomes of *wAnD* and *wAnM* [37], whole-genome sequences of 15 *Wolbachia* genomes were downloaded from the National Center for Biotechnology Information (NCBI), with these genomes spanning supergroups A through to F (full information available in Table S1, available with the online version of this article). An additional 25 *Wolbachia* genomes were also downloaded from the European Nucleotide Archive (ENA). These additional genomes were sequenced as part of a large-scale study [48] that looked at assembling *Wolbachia* genomes from a variety of existing sequencing data of various insects. All genome accession numbers used in this study, as well as a summary of their annotations used in this study, are provided in Table S1.

To confirm genome completeness, nucleotide sequences of all downloaded genomes, as well as the assembled genomes of *wAnD* and *wAnM*, were used as input into the program BUSCO (v5.0.0) [49], with the lineage option set to 'rickettsiales\_odb10'. This program analyses genome completeness via comparison against a selection of marker genes (364 genes in total) predicted to be present in single copies based on the input genome's lineage. Genomes that showed significantly lower completeness levels (less than 80% completeness) were excluded from orthologue and pathway analyses. This resulted in six of the *Wolbachia* genomes [48] being removed from these additional analyses.

### Phylogenetic, pangenome and metabolic pathway analysis

A total of 36 *Wolbachia* genomes were used for phylogenetic analysis of supergroup B *Wolbachia* specifically (genomes of *wAnD* and *wAnM*, 9 *Wolbachia* genomes from the NCBI and 25 from ENA). These genome sequences were used as input into the program wgsim (version 1.9) [50, 51], which 'shreds' the genomic template to generate genome fragments similar to sequencing reads. Base error, mutation, fraction of indels and indel extension probability were set to zero, read lengths set to 100 and a total of ten million reads simulated for each genome. These genome reads were then used to generate a single nucleotide variant (SNV) alignment via Snippy v4.6.0 [52] using the *wNo* genome as reference (genome accession no. GCA\_000376585.1). Gubbins v3.0.0 [53] was used for removing recombinant events. Recombination-free alignment of all 34 genomes was then analysed with IQ-TREE v1.6.12 [54] using default parameters, with a GTR substitution model using 1000 non-parametric bootstrap replicates for branch support.

This initial tree was further validated using single-copy orthologous genes identified by OrthoFinder v2.5.1 [55] (see relevant section in Methods). All single-copy orthologous genes that were identified from each *Wolbachia* strain were aligned using MAFFT v7.455 [56], before being concatenated together using SeqKit's concat v0.15.0 program [57], with the resultant alignment used as input into IQ-TREE v1.6.12 [54] using default parameters. The substitution model used was HIVw +F+R2, identified as the best-fit model by the ModelFinder program [58], and 1000 non-parametric bootstrap replicates to determine branch support.

### Orthologous group detection

Orthologous group detection was performed in two separate parts – the first was to compare protein-encoding sequences amongst *Wolbachia* of supergroup A through to F, whilst the second was to compare protein-encoding sequences amongst *Wolbachia* of supergroup B specifically. For orthologue analysis amongst *Wolbachia* of supergroup A through to F, RefSeq protein annotations for the 15 genomes downloaded from NCBI were used, alongside RefSeq protein annotations for *wAnD* and *wAnM*. Protein sequences from these 17 genomes were used as input into the program OrthoFinder (v2.5.1) [55], using default parameters. Orthogroups that were common or unique between all 17 *Wolbachia* strains were subsequently plotted using the R program package UpSetR (v1.4.0) [59]. Additional querying of the data was then performed using the R program package 'ComplexHeatmaps' (v2.5.5) [60].

For orthologous group detection amongst a wider selection of supergroup B *Wolbachia*, a total of 27 *Wolbachia* genomes were used (6 from NCBI, 19 from ENA, alongside the assembled genomes of *wAnD* and *wAnM*) (Table S1). Annotations for all *Wolbachia* genomes were generated using a local installation of the NCBI PGAP (build5508 2021-07-01) [61], and were used as input into the program OrthoFinder (v2.5.1) [55] using default parameters. Orthogroups were again visualized using the R program UpSetR (v1.4.0) [59], with additional data querying performed using the R program ComplexHeatmaps (v2.5.5) [60]. Genes of interest identified within these orthogroups, e.g. those that were unique to particular genomes, were further analysed using the Pfam website's sequence search [62, 63] and NCBI's BLASTP [64]. Comparison of the identified nucleotide regions that had similarity to the *Osmia lignaria* gene XP\_034172187.1 was performed by BLASTN and BLASTX, with visualizations performed using Easyfig [65].

### Construction of metabolic pathways

The genomes of both *wAnD* and *wAnM* were submitted to the NCBI Prokaryotic Annotation Pipeline, with a GenBank flatfile being generated as a result. This flatfile was then downloaded, and used as input into BioCyc's Pathway Tools program (v24.0) [66] and Pathologic (v24.0) [67, 68]. Pathologic is able to assign protein function and pathways to annotated genes based on name and/or automated BLAST hits. To address proteins with 'ambiguous' function within metabolic pathways, all predicted protein-encoding genes of both *wAnD* and *wAnM* were submitted to the EggNOG online server, which allows for the automated transfer of functional annotations (v2.0.1) [69]. Predicted protein-encoding genes were also submitted to the KEGG Automatic Annotation Server (KAAS, last updated 3 April 2015) [70], as a second method for functional annotation. Any proteins identified by Pathologic as having an ambiguous function were then manually cross-checked with the outputs of EggNOG and KAAS, and enzyme code numbers assigned. This process was repeated for a selection of *Wolbachia* genomes from supergroup B (Table S1). Once this process was completed, Pathway Tools' Pathway Overview and Comparison options were then used to compare pathways between the different *Wolbachia* strains. A selection of these biosynthetic and transport pathways was then made, based on prior literature investigating their importance to the *Wolbachia*-host endosymbiotic relationship. Gene presence and absence within these pathways was then manually scored, and plotted out into a heatmap using R's Ggplot2 package [71].

### Characterization of CI factor genes

Phylogenetic placement of the three sets of *cif* gene pairs from both *wAnD* and *wAnM* was made following the methods of Martinez *et al.* [15]. Briefly, nucleotide sequences for CI factor (*cif*) A and B genes for all five monophyletic types were obtained from the supplementary materials of Martinez *et al.* [15]. Partially sequenced *cif* genes were discarded, and the nucleotide sequences for *cifA* and *cifB* genes were aligned separately using the program MAFFT. The separate alignments were then used as input into the online GBLOCKS server (v0.91b) [72], with default 'stringent' parameters to filter out weakly conserved regions of the alignment. Once filtering was done, the separate nucleotide sequence alignments were then concatenated using SeqKit (v0.15.0) [57], and used as input for PhyML (v3.0) [73], using the GTR GAMMA substitution model of evolution and 1000 bootstrap replicates. The outputted Newick formatted tree was then annotated using the GGTree package in R (v2.2.4) [74].

Due to concerns that the *cifB* gene from *wAnD* may be approaching pseudogenization, four uninterrupted homologues of the intact *cifB* gene from *wAnD* were identified from the generated phylogenetic tree. These were the *cifB* genes from *wAra* pair 2, *wBai* pair 2, *wHa* pair 2 and *wNik* pair 2. Gene sequences for these were obtained from the supplementary data of a published study [15], and used as input into the program PRANK (v170427) [75] for codon-based alignments using the options -codon and -translate. The resultant alignment was then manually examined for the presence of potential pseudogenization or frameshifts.

### Ankyrin, prophage and insertion sequence (IS) element detection

Ankyrin domains were detected using five HMMer profiles (ID numbers PF00023.31, PF12796.8, PF13606.7, PF13637.7, PF13857.7) [63]. These profiles were generated via first downloading associated alignment files from the Pfam protein database [62] as Stockholm formatted seed files. The HMMer suite (v3.1b2) [63] was then used to build HMM (hidden Markov model) profiles from these seed files. These profiles were then compared against the protein amino acid sequences annotated from *wAnD* and *wAnM* to identify any protein-encoding genes containing an ankyrin domain. This analysis was then repeated for a selection of *Wolbachia* genomes to allow for direct comparisons to be made.

Prophage sequences were identified within the genomes of *wAnD* and *wAnM* using the PHASTER web server [76]. Assembled contig sequences of both genomes were uploaded separately to the server, checking the option to note that the input consists of multiple separate contigs. In the case of *wAnD* where prophage regions were detected, results were downloaded and manual curation of the identified prophage regions was performed using the Artemis genome browser [77] to identify prophage genes overlapping these regions. Additional BLASTX searches were performed on neighbouring genes against phage WO<sub>VitA1</sub> sequences (GenBank nucleotide reference sequence HQ906662.1) to screen for genes that may be associated with prophage WO's EAM.

IS element detection was performed by submission of the two completed *wAnD* and *wAnM* genomes to the ISSaga online server (v1.0) [78] and results tables obtained. Manual curation of hits identified by ISSaga was then performed, with putative IS elements first filtered for high similarity and coverage BLAST searches against the ISFinder database. Predicted IS elements were further

compared against the assembled genome and automated annotations from PGAP, confirming that they have been annotated appropriately.

## RESULTS

### Currently assembled *Wolbachia* genomes are small in size but supported by high completeness scores

As *Wolbachia* is an obligate intracellular endosymbiont, it has a highly reduced genome and can only be isolated from infected host material, posing a challenge to obtain complete, uncontaminated genome sequences. The updated, circularized genome assemblies of *Wolbachia* of *Anopheles demeilloni* (*wAnD*) has a total length of 1 231 247 bp, while *Wolbachia* of *Anopheles moucheti* (*wAnM*) has a genome length of 1 121 812 bp. This assembly showed overall even coverage throughout the genome, although there were up to three regions per genome greater than 1 kbp that showed higher than average coverage, indicating potentially collapsed repetitive regions (Fig. S1). While the currently assembled genome sizes are smaller compared to other analysed *Wolbachia* strains that reside within insects (particularly *wAnM*), they are larger than the genomes of those found in filarial nematodes, which have a maximum size of 1.08 Mbp amongst those compared in our analysis (Tables 1 and Fig. S1). RefSeq annotation of both *wAnD* and *wAnM* genomes identified 1157 and 1082 protein-encoding genes and 122 and 80 pseudogenes, respectively (Table 1). For comparison, the *Wolbachia* strains of *Aedes albopictus* (*wAlbB*) and *Culex quinquefasciatus* (*wPip*) maintained 1180 and 1241 protein-encoding genes, respectively.

As an obligate intracellular endosymbiont that may have multiple strains infecting the same host, assessing *Wolbachia* genome completeness is important to ensure contaminating reads from different strains are not incorporated, and that the assembly does not have significant gaps. Despite the smaller number of protein-encoding genes and genome size, both *wAnD* and *wAnM* were noted to contain over 98% of essential single-copy genes as determined by the BUSCO program [49] (Table 1), with only *wAnM* predicted to have one duplicated gene, indicating that both their respective hosts are infected with only a single strain of *Wolbachia*. These figures are in line with previously published and complete *Wolbachia* genomes of strains found within insects (Table 1), with examples such as the *Wolbachia* strains of *Drosophila* flies (*wMel*, *wRi*, *wHa*, *wAu* and *wNo*) and mosquitoes (*wAlbB*, *wPip*), all having completeness scores ranging from 97.5 to 99.5%. Additional details on these genomes used for comparison and their associated publications are available in Table S1.

### Different *Anopheles* species show potentially independent *Wolbachia* acquisition events

Whole-genome phylogenetic analysis was performed to better understand how *wAnD* and *wAnM* may have been acquired by *Anopheles*, utilizing the most closely related genome of *Wolbachia* of *Drosophila simulans* strain Noumea (*wNo*) as a reference [37]. Using a total of 36 genomes of *Wolbachia* strains from supergroup B, a SNV alignment of 2824 bp was generated. The midpoint-rooted tree of the SNV alignment (Fig. 1, unrooted radial tree shown in Fig. S2) placed both *wAnD* and *wAnM* within a clade that also includes *wNo*, and several *Wolbachia* strains that infect *Drosophila mauritiana* [48, 79, 80]. We observed a significant number of differences in this alignment between *wAnD* and *wAnM* strains, with a total of 824 SNVs between the two *Anopheles*-derived strains. By contrast, *wAnM* was shown to have a total of 408 and 417 SNVs shared between it and the *Wolbachia* strains of *Drosophila mauritiana* and *wNo*, respectively, suggesting that *wAnM* is more closely related to these strains than to *wAnD*. As an additional verification of results, a phylogenetic tree was also reconstructed utilizing 172 single-copy orthologues identified by the program OrthoFinder [55]. This generated a similar tree to that obtained previously, with *wAnM* and *wAnD* being members of a clade with *wNo* and *wMa/wMau*, and being distinct from known *Wolbachia* strains that infect mosquitoes, *wAlbB* and *wPip*. This phylogenetic tree is shown in Fig. S3.

Regardless of the method used to reconstruct the phylogenetic tree, it was observed that *wAlbB* and *wPip*, two known *Wolbachia* strains of mosquitoes, do not cluster together, and appear in clades separate from both *wAnM* and *wAnD*. This lack of host phylogenetic congruence can be seen throughout both generated trees (Figs 1 and S3), with Insecta host members from different orders appearing throughout. Exceptions to this observation come from *Wolbachia* genomes that have been sequenced from the same host, e.g. *Diaphorina citri* or *Drosophila mauritiana*. Such observations are similar to those in previous studies that predict how *Wolbachia* is not solely restricted to vertical transmission [81, 82] and could be an indication of independent horizontal acquisition of *wAnD* and *wAnM* in their current hosts, rather than an ancestral infection that has since been lost in other anopheline mosquitoes. Phylogenetic analysis of COII and ITS2 sequences of *Anopheles demeilloni* and *Anopheles moucheti* had previously indicated significant phylogenetic distances from both the *Anopheles gambiae* and *Anopheles funestus* complexes [37]. Furthermore, this study also provided no evidence of resident *Wolbachia* strains within *Anopheles marshallii*, a mosquito species closely related to *Anopheles demeilloni* and *Anopheles moucheti* [37].

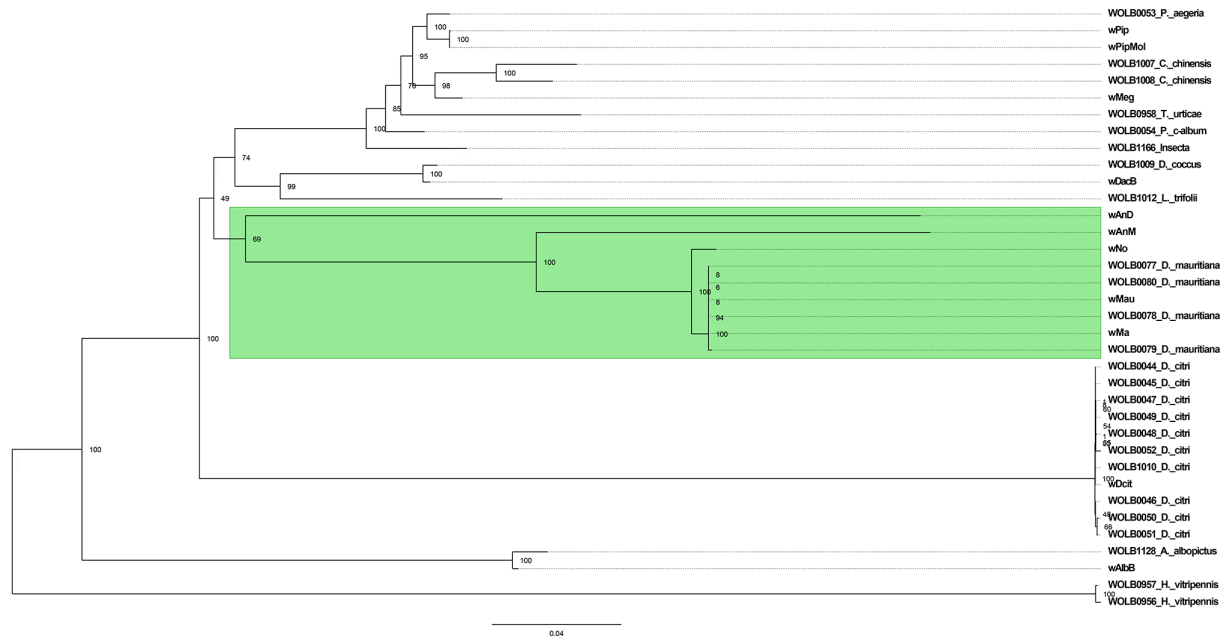
### *Wolbachia* core genome is conserved in *wAnM* and *wAnD* orthogroup analysis

Orthologous gene groups are important to identify in *Wolbachia* strains due to their wide distribution across supergroups and diverse hosts, whilst offering insights into the presence/absence of unique pathways that may be involved in host–bacterial symbiosis. For this, we compared the RefSeq annotations of *wAnD* and *wAnM* genomes against 17 *Wolbachia* genomes (Table 1).

**Table 1.** Summary table of a selection of different near-complete *Wolbachia* genomes and their general genome properties

Note the genomes of wAnD and wAnM (highlighted in bold text) have similar genome properties compared to other *Wolbachia* genomes, but have a relatively lower number of protein-encoding genes when compared against *Wolbachia* strains of supergroups A and B. Busco scores were calculated using the *Rickettsiales\_odb10* lineage, created on 06/03/2020 with a marker gene list total of 364.

Strain name	Host organism	Supergroup	Size (Mb)	G+C (mol%)	Total genes	No. of proteins	No. of pseudogenes	tRNA	rRNA	Other RNAs	No. of ankyrin proteins	busco score (out of 364)
wAu	<i>Drosophila simulans</i>	A	1.27	35.22	1265	1099	125	34	3	4	35	362 (99.5%)
wHa	<i>Drosophila simulans</i>	A	1.30	35.09	1242	1110	91	34	3	4	36	362 (99.5%)
wMel	<i>Drosophila melanogaster</i>	A	1.27	35.23	1247	1144	103	34	3	4	27	361 (99.2%)
wRi	<i>Drosophila simulans</i>	A	1.45	35.16	1340	1245	95	35	3	4	33	360 (98.9%)
wAlbB	<i>Aedes albopictus</i>	B	1.49	34.50	1442	1180	221	34	3	4	38	355 (97.5%)
wAnD	<b><i>Anopheles demelloni</i></b>	<b>B</b>	<b>1.23</b>	<b>33.58</b>	<b>1320</b>	<b>1157</b>	<b>122</b>	<b>34</b>	<b>3</b>	<b>4</b>	<b>55</b>	<b>360 (98.9%)</b>
wAnM	<b><i>Anopheles moucheti</i></b>	<b>B</b>	<b>1.12</b>	<b>33.59</b>	<b>1203</b>	<b>1082</b>	<b>80</b>	<b>34</b>	<b>3</b>	<b>4</b>	<b>37</b>	<b>360 (98.9%)</b>
wMa	<i>Drosophila mauritiana</i>	B	1.27	34.00	1196	1055	100	34	3	4	49	360 (98.9%)
wMau	<i>Drosophila mauritiana</i>	B	1.27	34.00	1194	1054	99	34	3	4	49	361 (99.2%)
wMeg	<i>Chrysomya megacephala</i>	B	1.38	33.95	1268	1116	111	34	3	4	52	363 (99.7%)
wNo	<i>Drosophila simulans</i>	B	1.30	34.01	1208	1062	105	34	3	4	53	363 (99.7%)
wPip	<i>Culex quinquefasciatus</i>	B	1.48	34.19	1385	1241	103	34	3	4	63	362 (99.5%)
wOo	<i>Onchoerca ochengi</i>	C	0.96	32.07	733	645	47	34	3	4	2	346 (95.1%)
wOv	<i>Onchoerca volvulus</i>	C	0.96	32.07	734	648	45	34	3	4	3	345 (94.8%)
wBm	<i>Brugia malayi</i>	D	1.08	34.18	1029	845	143	34	3	4	18	357 (98.1%)
wFol	<i>Folsomia candida</i>	E	1.80	34.35	1662	1541	79	35	3	4	94	362 (99.5%)
wCle	<i>Cimex lectularius</i>	F	1.25	36.25	1238	1023	174	34	3	4	42	356 (97.8%)

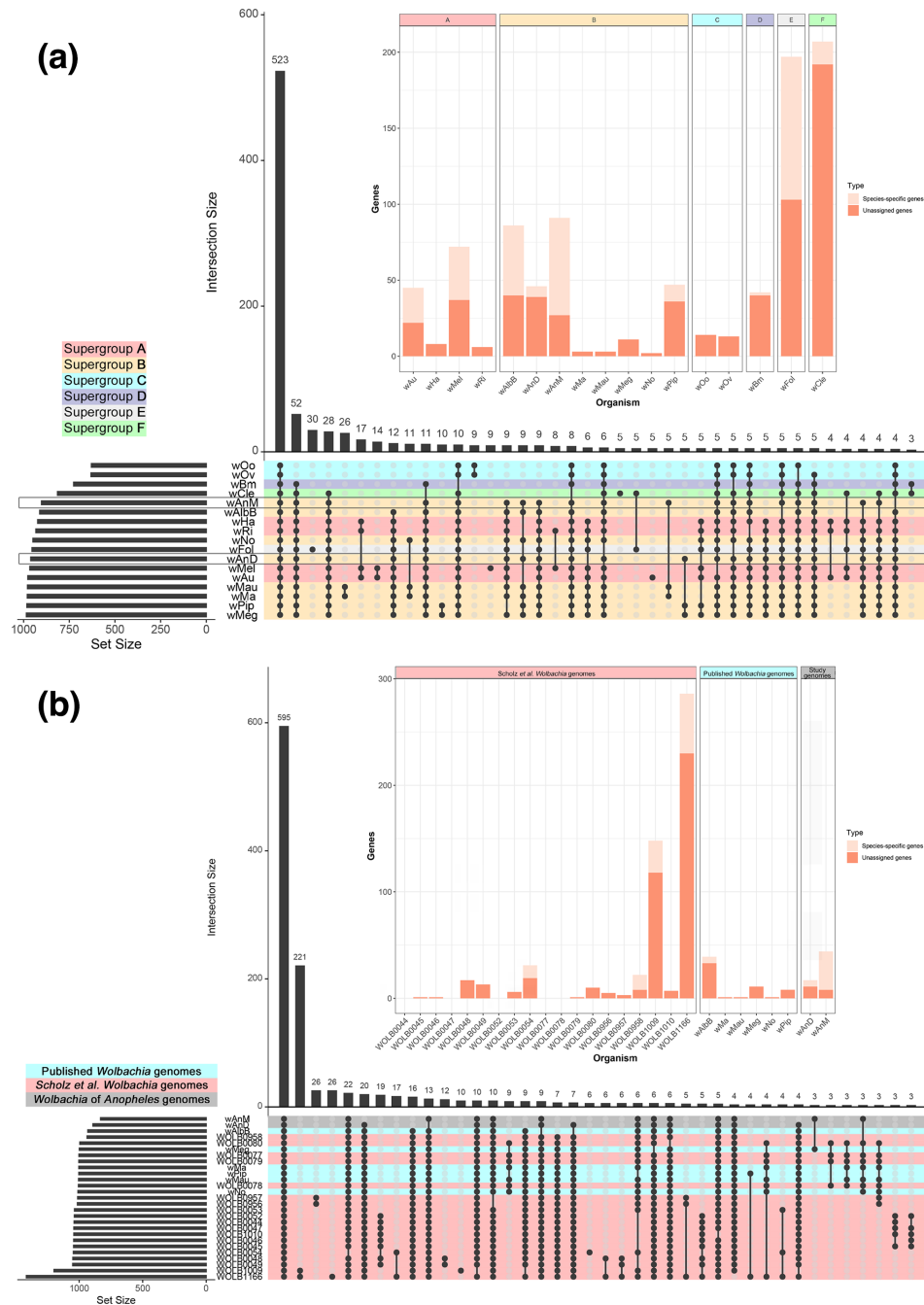


**Fig. 1.** Maximum-likelihood phylogenetic tree of whole-genome alignments of a selection of *Wolbachia* genomes, using 1000 bootstrap replicates. Genomes with names beginning with WOLB followed by four digits were assembled by Scholz *et al.* [49]. Other genomes, with the exception of wAnM and wAnD, are the results of previous sequencing efforts, with acronyms as described in Table 1. The tree is midpoint rooted. Note how wAnM and wAnD are present within a clade alongside wNo and several assembled genomes of *Wolbachia* from *Drosophila mauritiana* (green highlight). By contrast, previously sequenced *Wolbachia* of mosquitoes wPip/wPipMol and wAlbB are present in separate clades. Scale bar indicates relative branch lengths of the phylogenetic tree. An unrooted radial tree of this figure is included as Fig. S2.

A total of 18404 genes were analysed, with 96.8% of these assigned to 1300 orthogroups, and the remainder left unassigned to any orthogroup. Across the 17 *Wolbachia* strains analysed, a core genome of 9031 genes distributed across 523 orthogroups was identified (i.e. 40.2% of all identified orthogroups comprising 49.1% of total genes analysed can be considered as part of the core genome, defined as the genes and their protein products that are present in all analysed genomes), with 501 of these orthogroups containing single-copy genes. Outside of this core genome, the number of shared orthogroups is noticeably lower (Fig. 2a), and no orthogroups were unique to *Wolbachia* supergroup B strains. For wAnD, 39 genes were not assigned to an orthogroup, and one species-specific orthogroup (paralogues present in only one species) was identified containing seven genes (Fig. 2a, inset). By contrast, wAnM was noted to have 27 unassigned genes, as well as three species-specific orthogroups containing a total of 64 genes. None of the protein products for these genes had identifiable protein domains. Two orthogroups containing single-copy genes were identified that were specific to both wAnD and wAnM, although again none of these had identifiable protein domains.

Further comparisons were performed using a wider selection of *Wolbachia* supergroup B strains, including 6 *Wolbachia* genomes used in the previous analysis, as well as a further 19 draft genomes [48] with over 80% completeness. For consistency, all genomes were annotated using a local installation of NCBI's Prokaryotic Genome Annotation Pipeline (PGAP) [61]. In total, 32,064 genes annotated across the 27 genomes were used, of which 98.4% were assigned to 1,669 orthogroups (Fig. 2b). A core genome (genes and their protein products that are present in all analysed genomes) was identified containing 16,957 genes distributed across 595 orthogroups (47.6% of total genes were assigned to 36.8% of all orthogroups). Of these 595 orthogroups, 172 contain single-copy genes. A total of 11 and eight genes were not assigned to an orthogroup for wAnD and wAnM, respectively. One species-specific orthogroup was identified in both wAnD and wAnM (containing six and 36 genes respectively). Similar to the previous comparison, one orthogroup was identified as specific to both wAnD and wAnM, containing single-copy orthologues from both genomes that did not have any identifiable protein domains.

It was interesting to see that the number of orthogroups that could be considered as part of the core genome is less than 50% for both comparisons conducted here. We observed a total of 90 orthogroups that are not considered core due to their absence within *Wolbachia* strains of filarial nematodes from supergroups C and/or D specifically, whilst supergroup F strains has 30 unique orthogroups. Additionally, the genomes of *Wolbachia* from *Drosophila mauritiana* (wMa and wMau in Fig. 2a) shared 26 unique orthogroups. This observation of an extensive accessory genome has been reported in the past, even among closely related *Wolbachia* strains [83].



**Fig. 2.** Overview of identified orthogroups amongst *Wolbachia*. (a) Graphical representation of set notation of 17 near-complete *Wolbachia* genomes from six of the main supergroups using UpSetR, and the protein orthologues that they encode. Each genome (one per row at the bottom half of the image) is treated as a ‘set’ containing a certain number of orthogroups (denoted by the bar graph on the bottom left of the image). The various permutations of intersections are denoted by the ball-and-stick diagram at the bottom of the image, and the size of these intersections denoted by the bar graph at the top of the image. *Wolbachia* genomes are colour-coded based on their supergroup organization, and the genomes of wAnD and wAnM are highlighted by an additional grey outline on their row. Note how an intersect of all 17 *Wolbachia* genomes was identified as containing the vast majority of orthogroups – a core proteome total of 523 orthogroups (first bar from left). All other subsequent permutations of intersects contain less than 53 orthogroups. There were no intersects that uniquely contained only supergroup B *Wolbachia*. The inset stacked bar chart shows the distribution of singleton (i.e. genes that do not belong to an orthogroup, dark orange/red bar segment) and strain-specific orthogroups (i.e. genes that belong to an orthogroup unique to that *Wolbachia* strain, light orange/red segment). (b) Graphical representation of set notation of 27 supergroup B *Wolbachia* genomes, using the format as described for (a). *Wolbachia* genomes include 6 existing published complete genomes (red highlight), and 19 recently assembled genomes from Scholz *et al.* [49] (blue highlight), alongside the 2 recently assembled genomes wAnM and wAnD (grey highlight). Analysis was performed on local PGAP annotations of all 27 *Wolbachia* genomes. Note how the intersect of all 27 *Wolbachia* genomes shows a core proteome of 595 orthogroups, with the second largest intersect containing 221 orthogroups shared between the Scholz *et al.* [49] genome assembly for *Wolbachia* of *Dactylopius coccus* and an unidentified Insecta.



## Despite smaller genomes, the *Wolbachia* spp. core metabolic pathways are conserved in *wAnD* and *wAnM*

Orthogroup analysis of *wAnD* and *wAnM* indicated a high degree of conservation of supergroup B metabolic capacity, and to confirm this KAAS [70] was used to assign KEGG orthology. A total of 677 and 660 protein-encoding genes were assigned a KEGG orthologue (KO) number for *wAnD* and *wAnM*, respectively. Subsequent visualization and manual annotation identified complete biosynthetic pathways that have previously been considered of interest with respect to *Wolbachia*–host symbiosis (Fig. 3a). This includes pathways for riboflavin, purines, pyrimidines and haem biosynthesis; and showed all pathways as present in other supergroup B isolates' genomes. Additionally, both *wAnD* and *wAnM* also contained a suite of metabolite transport and secretion systems common to other *Wolbachia* strains that includes haem, zinc, iron (III), lipoproteins and phospholipids (Fig. 3a). This conservation of pathways was also observed when the analysis was focused on only *Wolbachia* from supergroup B strains [48] (Fig. 3b). In addition to these biosynthetic pathways, the T4SSs and Sec-secretion systems were also maintained in both *Wolbachia* genomes. The T4SSs are known to play roles in infection and survival for a diverse range of symbiotic and pathogenic intracellular bacteria [84, 85]. Both *Wolbachia* genomes contained a total of 15 T4SS-related genes, organized into two operonic regions and four individual genes spread across the genome. *Wolbachia* strains in the filarial nematode *Brugia malayi* has been predicted to utilize its T4SS to secrete protein effector molecules to avoid autophagy pathways and aid in actin cytoskeleton reformation, allowing intracellular mobility [86]. Such processes may also be conserved within *Wolbachia* strains residing within insects, such as *wAnD* and *wAnM*.

## Prophage WO regions are absent, or highly degenerated, in *wAnM* and *wAnD*

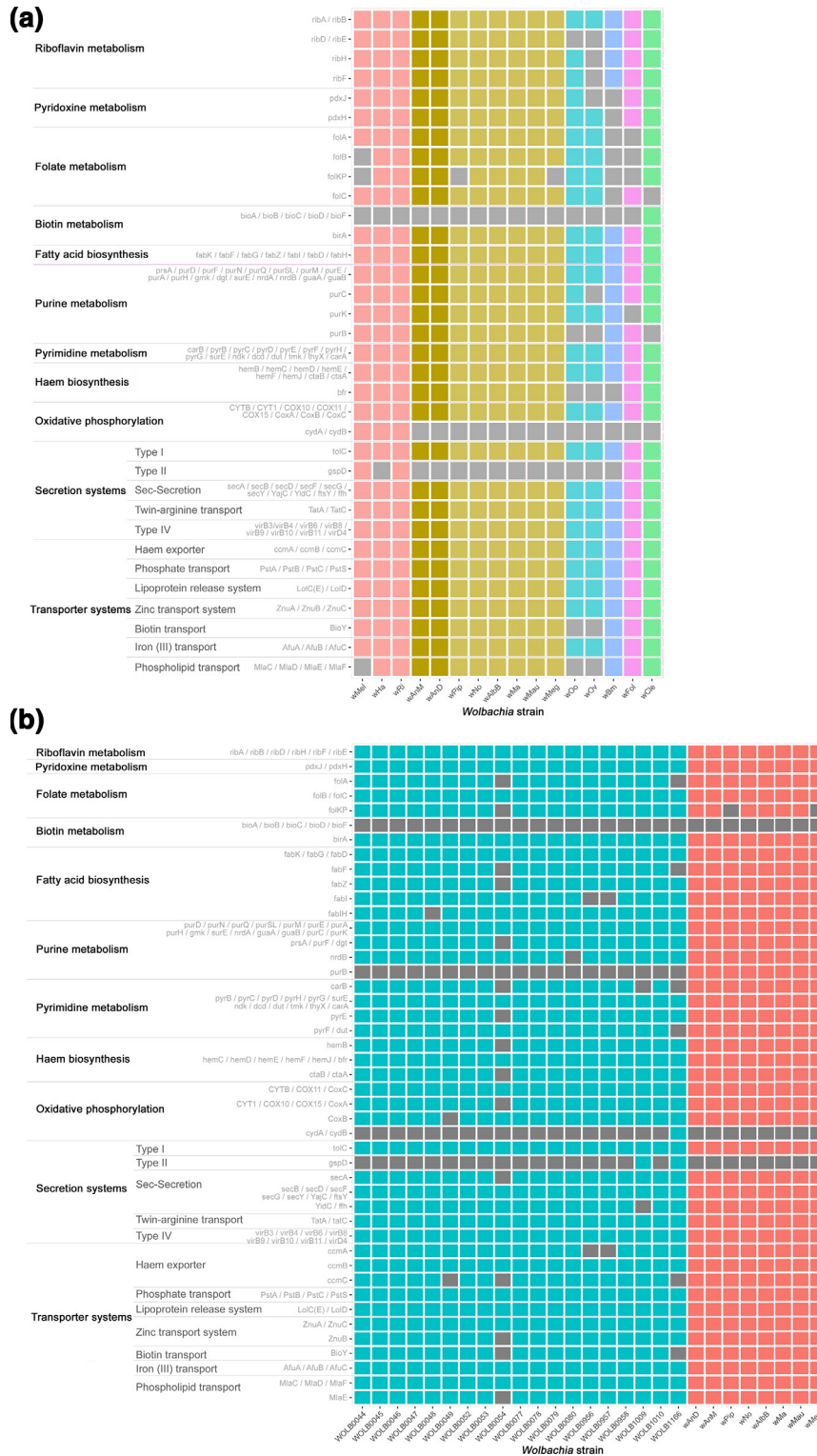
*Wolbachia* strains found within insects are frequently infected by a bacteriophage known as phage WO [87], with prophage sequences predicted to be common in the genomes of *Wolbachia* strains of insects [13, 88]. These prophage regions are known to maintain an EAM [89], a group of genes that encode protein domains homologous to those found in eukaryotes. This has resulted in predictions that these genes influence host–*Wolbachia* interactions by mimicking and interacting with host proteins [89]. Additionally, genes that have been implicated in the mode of action for CI have typically been found localized within these prophage EAM regions [11, 12, 16, 89].

In contrast to other *Wolbachia* strains that reside within mosquitoes, *wAnM* contained no prophage fragments identifiable via the PHASTER web server. To confirm this, we aligned the genomes of both *Wolbachia* strains from the two *Anopheles* species to their closest relative, *wNo* from *Drosophila simulans*. The *Wolbachia* genome of *wNo* was previously observed to have four prophage-like regions [90], ranging in size from 5.7 to 47.2 kbp. Initial comparisons of the genomes showed notable gaps within the *wAnM* genome when compared to *wNo*, although the same regions appear partially present in *wAnD* (Fig. 4a, b). Overlaying coordinates for the four prophage regions that were known to be present in *wNo* [90] onto this comparison, it was observed that the gaps in alignment with *wAnM* were centred on these *wNo* prophage regions (Fig. 4a). When the original sequencing reads were mapped to the *wNo* genome, we observe very low read coverage on *wNo* prophage segments (Fig. 2), whilst these reads showed even coverage of the *wAnM* genome. This indicates that these prophage regions are absent in the current assembly of *wAnM*.

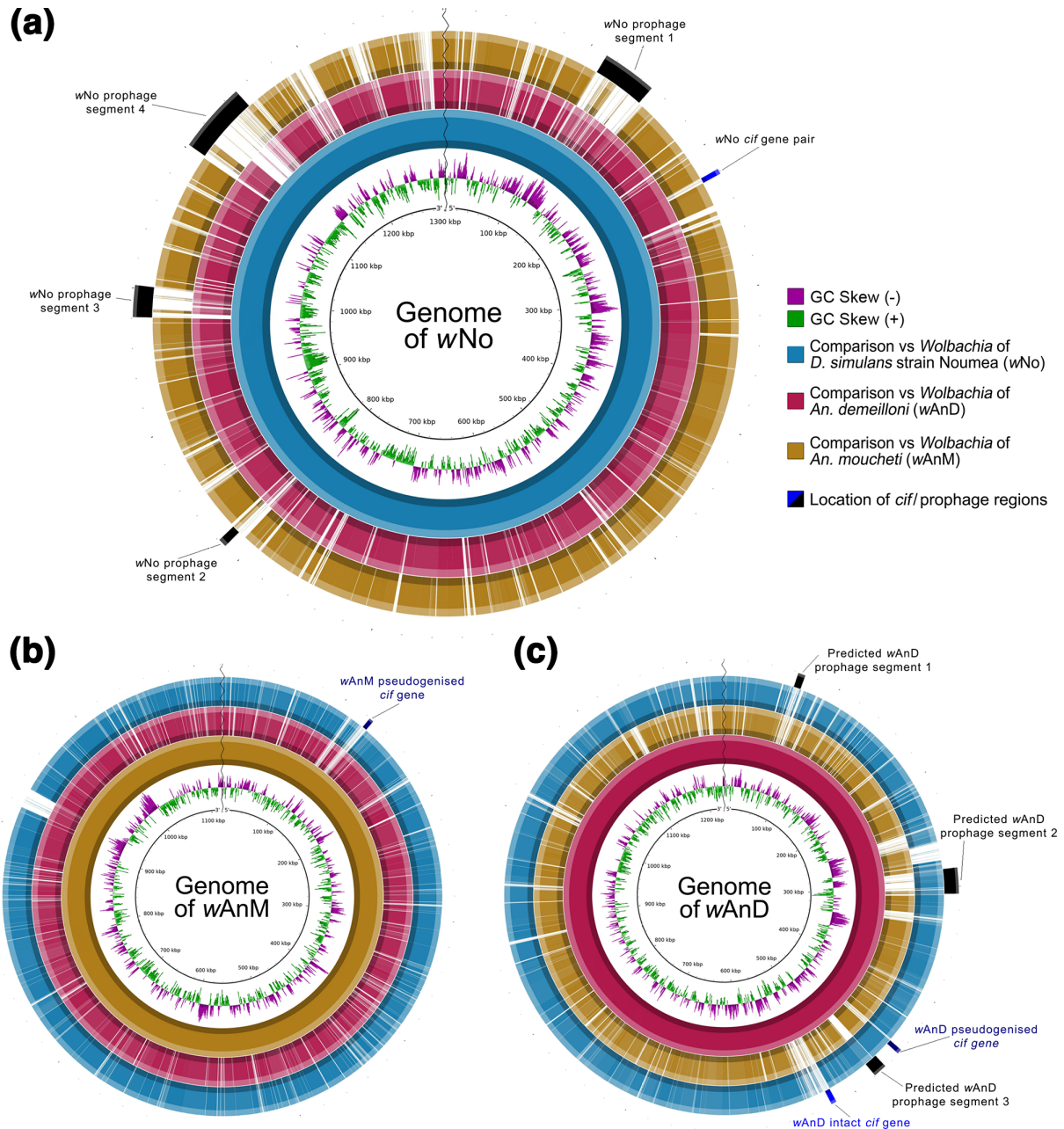
Within *wAnD*, analysis via the PHASTER web server and subsequent BLASTX searches of surrounding regions identified two prophage fragments of lengths 6.3 and 22.1 kbp. BLASTX searches also identified an additional prophage-like region of length 11.6 kbp (Fig. 4c). The total length of these prophage fragments (approx. 40 kbp) is shorter than published phage WO genomes (lengths of between 55 and 65 kbp) [89, 91]. The two prophage regions identified by PHASTER are predicted to encode a total of 50 genes, 16 of which were predicted to be interrupted by either stop codons or frameshifts. The prophage-like region identified after manual curation contained 13 genes, of which 7 were predicted to be interrupted. Two of these three regions contained structural phage genes that were either intact or interrupted, with examples including phage tail, baseplate, head–tail connectors and capsid proteins (Fig. S4). These observations, specifically the large number of interrupted structural phage genes identified in *wAnD*, indicates that this *Wolbachia* strain likely maintains a cryptic prophage, incapable of producing active phage particles.

## CI factors are conserved in *wAnM* and *wAnD*

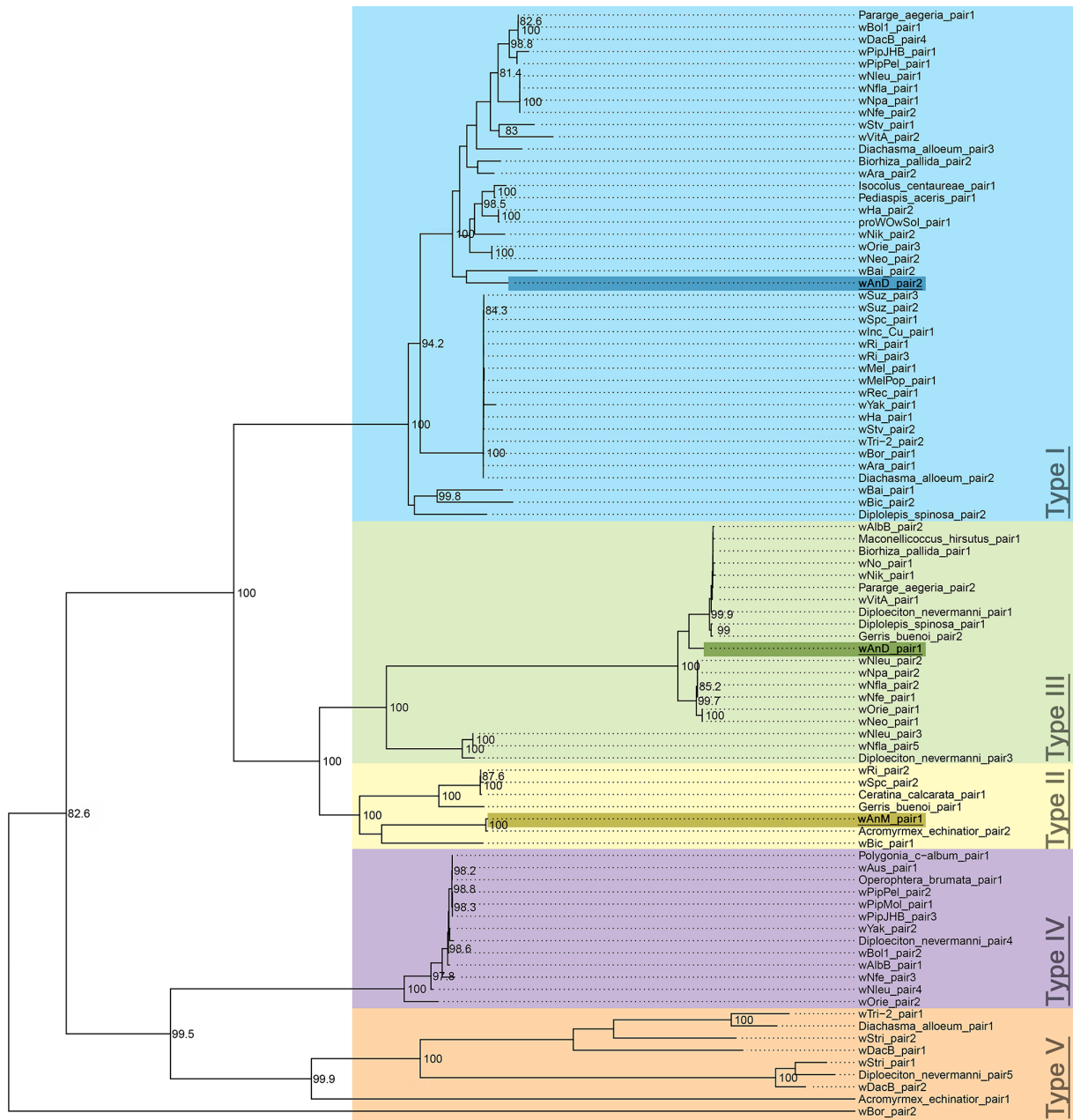
We previously reported that the genome of *wAnD* contains one intact pair of *cif* genes (JSQ73\_02850, JSQ73\_02855), and a second pair that showed interruptions in both genes (JSQ73\_02500 through to JSQ73\_02515) [37]. In turn, the genome of *wAnM* contains one pair of *cif* genes, although two internal stop codons were identified within *cifB* [37]. Phylogenetic analysis of the concatenated nucleotide sequences of *cifA* and *cifB* identified *wAnD*'s intact *cif* gene pair as clustering with the type I group, and its pseudogenized pair clustering with the type III group, in line with previous observations [15]. Further investigation of the intact *cifB* gene in *wAnD* via a codon-based alignment identified no evidence for potential pseudogenization events (Fig. S5). In comparison, the *cif* gene pair of *wAnM* clusters with the type II group (IYZ83\_00740 through to IYZ83\_00755; Fig. 5, unrooted radial tree shown in Fig. S6).



**Fig. 3.** Heatmap representation of the presence–absence of various genes in metabolic and secretion/transport system pathways amongst a selection of *Wolbachia* genomes. The analysed genomes are arrayed on the x-axis, with colours of the heatmap representing the various analysed supergroups. The y-axis in turn represents different genes and metabolic pathways of interest to *Wolbachia* studies. Greys in the heatmap represent an absence of a gene within the respective genome. (a) Comparative illustration of 16 near-complete *Wolbachia* genomes. Column colours are based on *Wolbachia* supergroup using a similar scheme to Fig. 2, with columns representing the genomes of wAnD and wAnM highlighted with a more intense colour of the heatmap. These genomes were observed to maintain all the genes and pathways common to supergroup B *Wolbachia*. (b) Comparative illustration of 29 *Wolbachia* genomes of supergroup B. Columns are coloured based on their origin, with blue columns being genomes from the study by Scholz et al. [49], and red being genomes from existing *Wolbachia*, including wAnD and wAnM.



**Fig. 4.** BLAST Ring Image Generator (BRIG) visualization of prophage regions in the genomes of wNo, wAnM and wAnD when compared to one another. Each individual ring represents the presence or absence of similarity for a specific genome (represented by the different colours, see the key on the right of a) against a template genome (represented by the innermost solid colour ring, and the name at the centre of the panel). Presence of similarity is defined as the query genome (the ring) having greater than 50% similarity to the template genome. The outermost ring for each figure part contains information on predicted prophage and *cif* gene localizations. (a) Comparison of wAnM and wAnD against a wNo template genome (1301823bp in length). Note how the black bars representing predicted prophage regions as described by Ellegaard *et al.* [91] overlap areas with no similarity against the wAnM genome, whilst having some similarity to the wAnD genome. Also, note how its single, intact *cif* gene pair is located separately from previously predicted prophage regions. (b) Comparison against a wAnM template genome (1121812bp in length). Note how this genome was reported to contain no prophage regions, and its single pseudogenized *cif* gene pair is located in an area with no similarity to both wNo and wAnD. (c) Comparison against a wAnD template genome (1231247bp in length). Predicted prophage segments 1 and 2 were predicted by the PHASTER web server, with segment 3 predicted by BLASTX searches against the prophage regions WOVitA1 and WOCauB1 through to B3, as identified by Bordenstein *et al.* [90]. Note how of the three predicted prophage regions, two showed similarity to the wNo genome, and one showed no similarity to either genome. Also, note how its two *cif* gene pairs are located separate from, but close to, predicted prophage segment 3. In addition, note how the intact *cif* gene pair appears within a region that shows weak to no similarity against both wAnM and wNo.



**Fig. 5.** Maximum-likelihood phylogenetic tree of concatenated *cif* gene nucleotide alignments, built following the methods of Martinez *et al.* [16] with 1000 bootstrap replicates. Only bootstrap values of over 80% are shown. The five types of concatenated *cif* genes are highlighted with different colours, and their corresponding types annotated. The tree is midpoint rooted. The two pairs of *cif* genes of *wAnD* were previously noted to be members of type I and type III, which is confirmed by this repeated analysis. The pair of *cif* genes in *wAnM* can be found in the well-supported type II clade. An unrooted radial tree of this figure is included as Fig. S6.

Within *wAnD*, the interrupted *cif* genes were a combined 3.6 kbp in length and were located upstream of one of the prophage regions identified by the PHASTER web server (Fig. 4c). Following this, the intact *cif* genes of *wAnD* combined measured 6.0 kbp in length, and were approximately 69.5 kbp downstream of the interrupted *cif* genes (Fig. 4c). By contrast, the interrupted *cif* genes of *wAnM* were of a combined 3.6 kbp in length (Fig. 4b). Interestingly, none of the three identified pairs of *cif* genes within *wAnD* and *wAnM* were located directly next to or within prophage regions, although the two within *wAnD* are located close to one (Fig. 4c). This is similar to *wNo*, whose single intact pair of *cif* genes was observed to be separate from predicted prophage regions (Fig. 4a). Separately, it should be noted that the *cif* gene pair of *wAnM* appears to be a unique insertion that is also not present in *wAnD* (Fig. 4b).

## DISCUSSION

This study provides a comprehensive analysis of two *Wolbachia* strains recently identified within *Anopheles* mosquitoes. Their high density and prevalence rates within field populations provides an opportunity to better understand *Wolbachia*–host interactions, as well as providing a potential tool to aid in interrupting the spread of *Plasmodium* parasites. One of the first observations from this study is that the *Anopheles*-infecting *Wolbachia* strains are not monophyletic with other *Wolbachia* strains from mosquitoes (*wAlbB* and *wPip*). Instead, both *wAnD* and *wAnM* were located within a clade that includes several *Wolbachia* strains found within *Drosophila simulans* and *Drosophila mauritiana*. There have been multiple studies that show horizontal transmission of *Wolbachia* occurs regularly [81, 82], and is even possible via a plant intermediate [92]. This potential for horizontal transmission is further emphasized by a recent survey that assembled over 1000 *Wolbachia* genomes from existing sequence data [48]. These genome assemblies are primarily distributed across various *Wolbachia* strains from supergroups A and B, whilst also generating multiple *Wolbachia* assemblies from the same host [48]. This study observed how closely related *Wolbachia* strains can be found in taxonomically unrelated hosts, as well as finding no meaningful phylogenetic clustering of different hosts and their corresponding resident *Wolbachia* strains. Such observations are similar with what is observed here with the whole-genome phylogeny of *wAnD* and *wAnM*, in relation to the wider supergroup B strains and their insect hosts.

When compared against these other sequenced *Wolbachia* strains, analysis of the *wAnD* and *wAnM* strains indicates that they maintain relatively small genome sizes for strains found within insects. This may change in the future as the short-read sequencing assembly used in this study is unable to resolve potential repetitive regions, as observed via comparison of read depths against the genomes. Despite their reduced sizes, both the *wAnD* and *wAnM* strains maintain similar metabolic and transport pathways found in other *Wolbachia* strains. Additionally, no biosynthetic pathways were identified that could indicate a previously unknown feature acquired in these two strains found in *Anopheles* mosquitoes. Known pathways of relevance for *Wolbachia* include haem and nucleotide biosynthetic pathways [93], as well as transport components such as the T4SS for secreting potential protein effectors [84]. The observation of smaller genome sizes could be attributed to a reduced number of mobile elements, specifically prophage regions, when compared to other *Wolbachia* strains that reside in mosquitoes, such as *wAlbB* [94] and *wPip* [95].

Following on from this, it is interesting to see how *wAnD* has degenerated, and likely cryptic, prophage regions in comparison to its closest relative *wNo*, whilst *wAnM* lacks prophage regions entirely. Prophage regions in *Wolbachia* genomes are known to vary significantly, with some strains maintaining duplicate prophage insertions that can encode a functional prophage [90, 95], whilst others have been found to be degenerated [90, 94, 96]. Having said this, the complete lack of prophage regions seen in *wAnM* has so far only been reported within *Wolbachia* strains that infect nematodes [97, 98]. Furthermore, these *cif* genes were noted to be separate from any prophage regions, contrary to previous observations and expectations for these two features to be co-localized [11, 12, 25, 89, 99]. However, this separation of *cif* genes is not unique to just these two *Wolbachia* strains in *Anopheles*, but is also true in the closely related *Wolbachia* strains *wNo*, *wMa* and *wMau* (the first infecting *Drosophila simulans*, the latter two *Drosophila mauritiana*), which have been shown to maintain *cif* genes that are distinctly separate from any prophage WO region [79, 100] (Fig. S7). It is tempting to speculate that this separation of *cif* genes and prophage regions may be a unique feature of this clade of *Wolbachia*. For comparison, the genomes of both *wAlbB* and *wPip* (*Wolbachia* strains found in *Aedes* and *Culex* mosquitoes) maintain *cif* genes that are associated with prophage WO regions [11, 94]. In addition to this separation from prophage regions, both strains *wMa* and *wMau* were observed to have an interrupted *cifB* gene [79, 100], similar to what is observed in *wAnM*, and are both incapable of inducing CI, but capable of rescuing it, when crossed with *wNo*-infected mates [101, 102]. It should be noted that an interrupted *cifB* gene does not automatically mean an inability to induce CI, as it is known that several strains of *Wolbachia* that maintain intact *cifA* genes but interrupted *cifB* genes can induce a weak form of CI in their hosts [15, 103–105]. Further analysis to determine *wAnM*'s ability to cause CI will be necessary for further conclusions. One notable difference with the *cif* genes found in *wAnM* and *wAnD* is that all three of *wNo*, *wMa* and *wMau*'s *cif* gene pairs are found within the type III phylogenetic group, whereas the *cif* gene pair identified in *wAnM* can be placed within type II, which is unique amongst this clade of *Wolbachia*. BRIG comparisons of the different genomes appear to indicate this *cif* gene pair as a unique insertion. Furthermore, whilst *wAnD*'s degenerated *cif* gene pair was noted to be a member of the type III group, its intact *cif* gene pair also appears unique among this group of *Wolbachia* as a member of type I. Like *wAnM*'s sole *cif* gene pair, this intact *cif* gene appears to be a unique insertion event, separate from prophage elements.

How such insertion events within both *wAnD* and *wAnM* have come to happen, and where they have come from, is currently an open question that warrants further investigation, alongside how this group of *Wolbachia* maintain *cif* gene pairs that appear separate from identifiable prophage WO regions. One possible explanation is that the recent ancestors for these strains of *Wolbachia* may have acquired these *cif* genes from a recent phage WO insertion that has very recently become degenerated [100]. Alternatively, these prophage regions could have been removed from the genome by phage excision events. Previous publications have discussed what could happen to the *cif* gene pairs, as well as the *Wolbachia* that carry them, once CI is no longer able to induce evolutionary pressure on their hosts [106–109]. For instance, a recent survey of CI genes in *Wolbachia* predicted how, without evolutionary pressure, these CI genes would likely degrade over time, starting with *cifB*, the 'toxin' component of the phenotype, followed by *cifA*, the 'antidote' component [15]. Alternatively, it has also been suggested that the degradation of the *cif* genes may be related to the absence of prophage regions [15, 16], with the former being an adaptation used by the latter to spread

within *Wolbachia* populations. Thus, once the prophage regions are removed, it is predicted that the *cif* genes, and, thus, the CI phenotype, will have no evolutionary pressure to maintain themselves within *Wolbachia* [15, 16]. We observe this occurring to some degree in this study, with the dissociation of prophage regions from the *cif* genes, the interrupted type III pair observed in *wAnD*, and how *wAnM* carries interruptions in its type II *cifB* gene specifically. We see no evidence of such interruptions beginning to occur within the intact type I *cif* gene pair in *wAnD*. Once the phenotype these *Wolbachia* strains exert on their hosts can be properly elucidated, a longitudinal study on the *cif* genes within them is imperative. The results of such study could allow for further insights into *Wolbachia* biology and the evolution of the CI phenotype.

Despite the questions as to how this may have occurred, the observed similarities and differences between *wAnM* and its related strains *wMa*, *wMau* and *wNo* are intriguing, considering the high, but variable, prevalence rates of *wAnM* in field populations of *Anopheles moucheti* [37, 43]. This prevalence rate is a feature shared with *wAnD* [37, 43], which is more likely to be capable of inducing CI due to the presence of intact *cif* genes from the type I group, which *wMel* shares. For comparison, our previous work had shown the prevalence rates of *wAnM* to be between 17.5 and 75%, which is slightly lower than *wAnD* prevalence rates, shown to be between 38.7 and 100% [37]. Yet the ability for *Wolbachia* to persist in populations without inducing CI is known, as there are instances of *Wolbachia* that stably infect host populations without any overt reproductive parasitism phenotype [9, 110, 111]. Explanations for this have focused on *Wolbachia* providing some form of fitness benefit to their host. For instance, the *Wolbachia* strain *wAu* of supergroup A is capable of spreading through lab-based, uninfected host populations of *Drosophila simulans* without inducing CI [112]. This persistence of *wAu* could be linked to an ability to induce protection against viral infections [113], and it is tempting to speculate that *Wolbachia* may provide protection against pathogens of the mosquito. While such studies focus on *Wolbachia* of supergroup A, there has been some evidence that *wMau* of supergroup B may also confer a fitness benefit for their host via stimulating egg production [100, 114]. Taken together, this highlights the importance of further study regarding *wAnM*'s effect on their host, whether this be weak CI or some form of fitness benefit, such as benefits to reproduction or an ability to inhibit *Plasmodium* or viral infection.

The identification of natural *Wolbachia* infections in *Anopheles* shows promise for future control strategies of *Plasmodium* parasites. Whilst these strains show no pathways that are uniquely present or absent, they do exhibit unusual genomic arrangements with regards to the presence of prophage and *cif* genes. This has potential implications on their relationship with their respective anopheline hosts, potentially making them good candidates for transinfection into other medically relevant *Anopheles* species, such as *Anopheles gambiae* s.s. Further studies would be required to fully examine these *Wolbachia* strains and elucidate their predicted phenotypes of CI and pathogen blocking, both in the context of natural and artificial associations.

#### Funding information

E.H. and G.L.H. acknowledge support from Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/V011278/1. S.T. was supported by the Wellcome SEED award 217303/Z/19/Z to E.H. L.C. was supported by National Institute of Allergy and Infectious Diseases (NIAID) grant R01-AI116811. T.W. and C.L.J. were supported by a Sir Henry Dale Wellcome Trust/Royal Society fellowship awarded to T.W. (101285) (<https://wellcome.org> and <https://royalsociety.org>). T.W. was also supported by a Royal Society challenge grant (CHG\R1\170036). G.L.H. was also supported by the BBSRC (BB/T001240/1), a Royal Society Wolfson Fellowship (RSWF\R1\180013), the NIH (R21AI138074), the Engineering and Physical Sciences Research Council (EPSRC) (EP/V043811/1), the UKRI (20197 and 85336), and the NIH (NIHR2000907).

#### Author contributions

S.Q., designed methodology, conducted the investigation and formal analysis, designed visuals, and wrote the original draft. L.C., designed methodology, conducted the investigation and formal analysis, and wrote the original draft. C.L.J., conceptualized the study, conducted the investigation and resource collection, and wrote the original draft. S.T., designed methodology and software. T.W., conceptualized the study, conducted the investigation and resource collection, secured funding, and wrote the original draft. G.L.H., conceptualized and supervised the study, secured funding, and wrote the original draft. E.H., conceptualized and supervised the study, designed the methodology, and wrote the original draft. All authors have read and approved the final manuscript.

#### Conflicts of interest

The authors declare that there are no conflicts of interest.

#### References

1. Quek S, Cerdeira C, Jeffries CL, Tomlinson S, Walker T, et al. *Wolbachia* endosymbionts in two *Anopheles* species indicates independent acquisitions and lack of prophage elements. *Figshare* 2022. DOI: 10.6084/m9.figshare.19576432.
2. Werren JH, Baldo L, Clark ME. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* 2008;6:741–751.
3. Weinert LA, Araujo EV, Ahmed MZ, Welch JJ. The incidence of bacterial endosymbionts in terrestrial arthropods. *Proc R Soc B* 2015;282:20150249.
4. Sazama EJ, Bosch MJ, Shouldis CS, Ouellette SP, Wesner JS. Incidence of *Wolbachia* in aquatic insects. *Ecol Evol* 2017;7:1165–1169.
5. Taylor MJ, Bordenstein SR, Slatko B. Microbe Profile: *Wolbachia*: a sex selector, a viral protector and a target to treat filarial nematodes. *Microbiology* 2018;164:1345–1347.
6. Kaur R, Shropshire JD, Cross KL, Leigh B, Mansueto AJ, et al. Living in the endosymbiotic world of *Wolbachia*: a centennial review. *Cell Host Microbe* 2021;29:879–893.
7. Baldo L, Werren JH. Revisiting *Wolbachia* supergroup typing based on WSP: spurious lineages and discordance with MLST. *Curr Microbiol* 2007;55:81–87.
8. Baldo L, Dunning Hotopp JC, Jolley KA, Bordenstein SR, Biber SA, et al. Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Appl Environ Microbiol* 2006;72:7098–7110.

9. Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proc Natl Acad Sci* 2010;107:769–774.
10. Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, et al. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. *Proc Natl Acad Sci USA* 2014;111:10257–10262.
11. LePage DP, Metcalf JA, Bordenstein SR, On J, Perlmutter JI, et al. Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. *Nature* 2017;543:243–247.
12. Beckmann JF, Ronau JA, Hochstrasser M. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. *Nat Microbiol* 2017;2:17007.
13. Bordenstein SR, Wernegreen JJ. Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Mol Biol Evol* 2004;21:1981–1991.
14. Bordenstein SR, Brooks AW, Bordenstein SR, et al. Evolutionary genetics of cytoplasmic incompatibility genes *cifA* and *cifB* prophage WO *Wolbachia*. *Genome Biol Evol* 2018;10:434–451.
15. Martinez J, Klasson L, Welch JJ, Jiggins FM. Life and death of selfish genes: comparative genomics reveals the dynamic evolution of cytoplasmic incompatibility. *Mol Biol Evol* 2021;38:2–15.
16. Beckmann JF, Bonneau M, Chen H, Hochstrasser M, Poinot D, et al. The toxin-antidote model of cytoplasmic incompatibility: genetics and evolutionary implications. *Trends Genet* 2019;35:175–185.
17. Gillespie JJ, Driscoll TP, Verhoeve VI, Rahman MS, Macaluso KR, et al. A tangled web: origins of reproductive parasitism. *Genome Biol Evol* 2018;10:2292–2309.
18. Werren JH. Biology of *Wolbachia*. *Annu Rev Entomol* 1997;42:587–609.
19. Serbus LR, Casper-Lindley C, Landmann F, Sullivan W. The genetics and cell biology of *Wolbachia*-host interactions. *Annu Rev Genet* 2008;42:683–707.
20. McMeniman CJ, Lane RV, Cass BN, Fong AWC, Sidhu M, et al. Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti*. *Science* 2009;323:141–144.
21. Walker T, Johnson PH, Moreira LA, Iturbe-Ormaetxe I, Frentiu FD, et al. The wMel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. *Nature* 2011;476:450–453.
22. Hoffmann AA, Montgomery BL, Popovici J, Iturbe-Ormaetxe I, Johnson PH, et al. Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. *Nature* 2011;476:454–457.
23. Crawford JE, Clarke DW, Criswell V, Desnoyer M, Cornel D, et al. Efficient production of male *Wolbachia*-infected *Aedes aegypti* mosquitoes enables large-scale suppression of wild populations. *Nat Biotechnol* 2020;38:482–492.
24. Shropshire JD, On J, Layton EM, Zhou H, Bordenstein SR. One prophage WO gene rescues cytoplasmic incompatibility in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 2018;115:4987–4991.
25. Shropshire JD, Bordenstein SR. Two-by-one model of cytoplasmic incompatibility: synthetic recapitulation by transgenic expression of *cifA* and *cifB* in *Drosophila*. *PLoS Genet* 2019;15:e1008221.
26. McGraw EA, O'Neill SL. Beyond insecticides: new thinking on an ancient problem. *Nat Rev Microbiol* 2013;11:181–193.
27. Bourtzis K, Dobson SL, Xi Z, Rasgon JL, Calvitti M, et al. Harnessing mosquito-*Wolbachia* symbiosis for vector and disease control. *Acta Trop* 2014;132:S150–S163.
28. Johnson KN. The impact of *Wolbachia* on virus infection in mosquitoes. *Viruses* 2015;7:5705–5717.
29. Utarini A, Indriani C, Ahmad RA, Tantowijoyo W, Arguni E, et al. Efficacy of *Wolbachia*-infected mosquito deployments for the control of dengue. *N Engl J Med* 2021;384:2177–2186.
30. Indriani C, Tantowijoyo W, Rancès E, Andari B, Prabowo E, et al. Reduced dengue incidence following deployments of *Wolbachia*-infected *Aedes aegypti* in Yogyakarta, Indonesia: a quasi-experimental trial using controlled interrupted time series analysis. *Gates Open Res* 2020;4:50.
31. Hughes GL, Koga R, Xue P, Fukatsu T, Rasgon JL. *Wolbachia* infections are virulent and inhibit the human malaria parasite *Plasmodium falciparum* in *Anopheles gambiae*. *PLoS Pathog* 2011;7:3–10.
32. Bian G, Joshi D, Dong Y, Lu P, Zhou G, et al. *Wolbachia* invades *Anopheles stephensi* populations and induces refractoriness to *Plasmodium* infection. *Science* 2013;340:748–751.
33. Joshi D, Pan X, McFadden MJ, Bevins D, Liang X, et al. The maternally inheritable *Wolbachia* wAlbB induces refractoriness to *Plasmodium berghei* in *Anopheles stephensi*. *Front Microbiol* 2017;8:366.
34. Joshi D, McFadden MJ, Bevins D, Zhang F, Xi Z. *Wolbachia* strain wAlbB confers both fitness costs and benefit on *Anopheles stephensi*. *Parasit Vectors* 2014;7:336.
35. Walker T, Moreira LA. Can *Wolbachia* be used to control malaria? *Mem Inst Oswaldo Cruz* 2011;106:212–217.
36. Chrostek E, Gerth M. Is *Anopheles gambiae* a natural host of *Wolbachia*? *mMBio* 2019;10:e00784–19.
37. Walker T, Quek S, Jeffries CL, Bandibabone J, Dhokiya V, et al. Stable high-density and maternally inherited *Wolbachia* infections in *Anopheles moucheti* and *Anopheles demeilloni* mosquitoes. *Curr Biol* 2021;31:2310–2320.
38. Baldini F, Segata N, Pompon J, Marcenac P, Shaw WR, et al. Evidence of natural *Wolbachia* infections in field populations of *Anopheles gambiae*. *Nat Commun* 2014;5:3985.
39. Gomes FM, Hixson BL, Tyner MDW, Ramirez JL, Canepa GE, et al. Effect of naturally occurring *Wolbachia* in *Anopheles gambiae* s.l. mosquitoes from Mali on *Plasmodium falciparum* malaria transmission. *Proc Natl Acad Sci USA* 2017;114:12566–12571.
40. Niang EHA, Bassene H, Makoundou P, Fenollar F, Weill M, et al. First report of natural *Wolbachia* infection in wild *Anopheles funestus* population in Senegal. *Malar J* 2018;17:408.
41. Ayala D, Akone-Ella O, Rahola N, Kengne P, Ngangue MF, et al. Natural *Wolbachia* infections are common in the major malaria vectors in Central Africa. *Evol Appl* 2019;12:1583–1594.
42. Wong ML, Liew JWK, Wong WK, Pramasivan S, Mohamed Hassan N, et al. Natural *Wolbachia* infection in field-collected *Anopheles* and other mosquito species from Malaysia. *Parasites Vectors* 2020;13:414.
43. Jeffries CL, Lawrence GG, Golovko G, Kristan M, Orsborne J, et al. Novel *Wolbachia* strains in *Anopheles malaria* vectors from sub-Saharan Africa. *Wellcome Open Res* 2018;3:113.
44. Tsai IJ, Otto TD, Berriman M. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biol* 2010;11:R41.
45. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, et al. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol* 2018;14:e1005944.
46. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004;14:1394–1403.
47. Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, et al. Using tablet for visual exploration of second-generation sequencing data. *Brief Bioinform* 2013;14:193–202.
48. Scholz M, Albanese D, Tuohy K, Donati C, Segata N, et al. Large scale genome reconstructions illuminate *Wolbachia* evolution. *Nat Commun* 2020;11:5235.
49. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;31:3210–3212.
50. Li H. Wgsim for simulating sequence reads from a reference genome; 2011. <https://github.com/lh3/wgsim>
51. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.

52. Seeman T. Snippy: fast bacterial variant calling from NGS reads; 2015. <https://github.com/tseemann/snippy>
53. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015;43:e15.
54. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.
55. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 2015;16:157.
56. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
57. Shen W, Le S, Li Y, Hu F, Zou Q. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* 2016;11:e0163962.
58. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.
59. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 2017;33:2938–2940.
60. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016;32:2847–2849.
61. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 2016;44:6614–6624.
62. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, et al. The Pfam protein families database in 2019. *Nucleic Acids Res* 2019;47:D427–D432.
63. Howard Hughes Medical Institute. HMMER: biosequence analysis using profile hidden Markov models; 2013. <http://hmmer.org/>
64. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
65. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics* 2011;27:1009–1010.
66. Karp PD, Midford PE, Billington R, Kothari A, Krummenacker M, et al. Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. *Brief Bioinform* 2021;22:109–126.
67. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, et al. The MetaCyc database of metabolic pathways and enzymes – a 2019 update. *Nucleic Acids Res* 2020;48:D445–D453.
68. Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, et al. The BioCyc collection of microbial genomes and metabolic pathways. *Brief Bioinform* 2019;20:1085–1093.
69. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol* 2017;34:2115–2122.
70. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 2007;35:W182–W185.
71. Wickham H. *ggplot2: Elegant Graphics for Data Analysis* (<https://ggplot2.tidyverse.org>). New York: Springer-Verlag; 2016.
72. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 2007;56:564–577.
73. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.
74. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2017;8:28–36.
75. Löytynoja A. Phylogeny-aware alignment with PRANK. *Methods Mol Biol* 2014;1079:155–170.
76. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;44:W16–W21.
77. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 2012;28:464–469.
78. Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol* 2011;12:R30.
79. Baião GC, Janice J, Galinou M, Klasson L. Comparative genomics reveals factors associated with phenotypic expression of *Wolbachia*. *Genome Biol Evol* 2021;13:evab111.
80. Lefoulon E, Vaisman N, Frydman HM, Sun L, Volland L, et al. Large enriched fragment targeted sequencing (LEFT-SEQ) applied to capture of *Wolbachia* genomes. *Sci Rep* 2019;9:5939.
81. O'Neill SL, Giordano R, Colbert AM, Karr TL, Robertson HM. 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. *Proc Natl Acad Sci USA* 1992;89:2699–2702.
82. Werren JH, Zhang W, Guo LR. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proc Biol Sci* 1995;261:55–63.
83. Ishmael N, Hotopp JCD, Ioannidis P, Biber S, Sakamoto J, et al. Extensive genomic diversity of closely related *Wolbachia* strains. *Microbiology* 2009;155:2211–2222.
84. Rancès E, Voronin D, Tran-Van V, Mavingui P. Genetic and functional characterization of the type IV secretion system in *Wolbachia*. *J Bacteriol* 2008;190:5020–5030.
85. Christie PJ, Whitaker N, González-Rivera C. Mechanism and structure of the bacterial type IV secretion systems. *Biochim Biophys Acta* 2014;1843:1578–1591.
86. Carpinone EM, Li Z, Mills MK, Foltz C, Brannon ER, et al. Identification of putative effectors of the type IV secretion system from the *Wolbachia* endosymbiont of *Brugia malayi*. *PLoS One* 2018;13:e0204736.
87. Kent BN, Bordenstein SR. Phage WO of *Wolbachia*: lambda of the endosymbiont world. *Trends Microbiol* 2010;18:173–181.
88. Gavotte L, Henri H, Stouthamer R, Charif D, Charlat S, et al. A survey of the bacteriophage WO in the endosymbiotic bacteria *Wolbachia*. *Mol Biol Evol* 2007;24:427–435.
89. Bordenstein SR, Bordenstein SR. Eukaryotic association module in phage WO genomes from *Wolbachia*. *Nat Commun* 2016;7:13155.
90. Ellegaard KM, Klasson L, Näslund K, Bourtzis K, Andersson SGE. Comparative genomics of *Wolbachia* and the bacterial species concept. *PLoS Genet* 2013;9:e1003381.
91. Kupritz J, Martin J, Fischer K, Curtis KC, Fauver JR, et al. Isolation and characterization of a novel bacteriophage WO from *Allonemobius socius* crickets in Missouri. *PLoS One* 2021;16:e0250051.
92. Chrostek E, Pelz-Stelinski K, Hurst GDD, Hughes GL. Horizontal transmission of intracellular insect symbionts via plants. *Front Microbiol* 2017;8:2237.
93. Gill AC, Darby AC, Makepeace BL. Iron necessity: the secret of *Wolbachia*'s success? *PLoS Negl Trop Dis* 2014;8:e3224.
94. Sinha A, Li Z, Sun L, Carlow CKS, Cordaux R. Complete genome sequence of the *Wolbachia* wAlbB endosymbiont of *Aedes albopictus*. *Genome Biol Evol* 2019;11:706–720.
95. Klasson L, Walker T, Sebahia M, Sanders MJ, Quail MA, et al. Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. *Mol Biol Evol* 2008;25:1877–1887.



96. Miao YH, Xiao JH, Huang DW. Distribution and evolution of the bacteriophage WO and its antagonism with *Wolbachia*. *Front Microbiol* 2020;11:595629.
97. Foster J, Ganatra M, Kamal I, Ware J, Makarova K, et al. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol* 2005;3:e121.
98. Lefoulon E, Clark T, Guerrero R, Cañizales I, Cardenas-Callirgos JM, et al. Diminutive, degraded but dissimilar: *Wolbachia* genomes from filarial nematodes do not conform to a single paradigm. *Microb Genom* 2020;6:e000487.
99. Shropshire JD, Rosenberg R, Bordenstein SR. The impacts of cytoplasmic incompatibility factor (cifA and cifB) genetic variation on phenotypes. *Genetics* 2021;217:iyaa007.
100. Meany MK, Conner WR, Richter SV, Bailey JA, Turelli M, et al. Loss of cytoplasmic incompatibility and minimal fecundity effects explain relatively low *Wolbachia* frequencies in *Drosophila mauritiana*. *Evolution* 2019;73:1278–1295.
101. Bourtzis K, Dobson SL, Braig HR, O'Neill SL. Rescuing *Wolbachia* have been overlooked. *Nature* 1998;391:852–853.
102. Charlat S, Le Chat L, Merçot H. Characterization of non-cytoplasmic incompatibility inducing *Wolbachia* in two continental African populations of *Drosophila simulans*. *Heredity* 2003;90:49–55.
103. Shropshire JD, Leigh B, Bordenstein SR. Symbiont-mediated cytoplasmic incompatibility: what have we learned in 50 years. *Elife* 2020;9:e61989.
104. Cooper BS, Ginsberg PS, Turelli M, Matute DR. *Wolbachia* in the *Drosophila yakuba* complex: pervasive frequency variation and weak cytoplasmic incompatibility, but no apparent effect on reproductive isolation. *Genetics* 2017;205:333–351.
105. Shoemaker DD, Katju V, Jaenike J. *Wolbachia* and the evolution of reproductive isolation between *Drosophila recens* and *Drosophila subquinaria*. *Evolution* 1999;53:1157–1164.
106. Turelli M. Evolution of incompatibility-inducing microbes and their hosts. *Evolution* 1994;48:1500–1513.
107. Hurst LD, McVean GT. Clade selection, reversible evolution and the persistence of selfish elements: the evolutionary dynamics of cytoplasmic incompatibility. *Proc R Soc Lond B* 1997;263:97–104.
108. Koehncke A, Telschow A, Werren JH, Hammerstein P. Life and death of an influential passenger: *Wolbachia* and the evolution of CI-modifiers by their hosts. *PLoS One* 2009;4:e4425.
109. Haygood R, Turelli M. Evolution of incompatibility-inducing microbes in subdivided host populations. *Evolution* 2009;63:432–447.
110. Dedeine F, Boulétreau M, Vavre F. *Wolbachia* requirement for oogenesis: occurrence within the genus *Asobara* (Hymenoptera, Braconidae) and evidence for intraspecific variation in *A. tabida*. *Heredity (Edinb)* 2005;95:394–400.
111. Kriesner P, Hoffmann AA, Lee SF, Turelli M, Weeks AR. Rapid sequential spread of two *Wolbachia* variants in *Drosophila simulans*. *PLoS Pathog* 2013;9:e1003607.
112. Kriesner P, Hoffmann AA. Rapid spread of a *Wolbachia* infection that does not affect host reproduction in *Drosophila simulans* cage populations. *Evolution* 2018;72:1475–1487.
113. Ant TH, Herd CS, Geoghegan V, Hoffmann AA, Sinkins SP, et al. The *Wolbachia* strain wAu provides highly efficient virus transmission blocking in *Aedes aegypti*. *PLoS Pathog* 2018;14:e1006815.
114. Fast EM, Toomey ME, Panaram K, Desjardins D, Kolaczyk ED, et al. *Wolbachia* enhance *Drosophila* stem cell proliferation and target the germline stem cell niche. *Science* 2011;334:990–992.

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

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

Find out more and submit your article at [microbiologyresearch.org](https://microbiologyresearch.org).