ISME Communications, 2024, 4(1), ycae078
https://doi.org/10.1093/ismeco/ycae078
Advance access publication: 11 June 2024
Original Article

Wolbachia populations across organs of individual *Culex pipiens*: highly conserved intra-individual core pangenome with inter-individual polymorphisms

Blandine Trouche[1], Hans Schrieke[1], Olivier Duron[1], A. Murat Eren[2,3], Julie Reveillaud[1,*]

[1]IRD, MIVEGEC, University of Montpellier, INRAE, CNRS, 34394 Montpellier, France
[2]Marine Biological Laboratory, Woods Hole, MA 02543, United States
[3]Helmholtz Institute for Functional Marine Biodiversity at the University of Oldenburg, 26129 Oldenburg, Germany

*Corresponding author: Julie Reveillaud, UMR MIVEGEC, INRAe, Animal Health Department, Bâtiment Hortus, 361 rue Jean-Francois Breton BP 5095, 34196 Montpellier Cedex 05, France. Email: reveillaud.j@gmail.com

Abstract

Wolbachia is a maternally inherited intracellular bacterium that infects a wide range of arthropods including mosquitoes. The endosymbiont is widely used in biocontrol strategies due to its capacity to modulate arthropod reproduction and limit pathogen transmission. *Wolbachia* infections in *Culex* spp. are generally assumed to be monoclonal but the potential presence of genetically distinct *Wolbachia* subpopulations within and between individual organs has not been investigated using whole genome sequencing. Here we reconstructed *Wolbachia* genomes from ovary and midgut metagenomes of single naturally infected *Culex pipiens* mosquitoes from Southern France to investigate patterns of intra- and inter-individual differences across mosquito organs. Our analyses revealed a remarkable degree of intra-individual conservancy among *Wolbachia* genomes from distinct organs of the same mosquito both at the level of gene presence–absence signal and single-nucleotide polymorphisms (SNPs). Yet, we identified several synonymous and non-synonymous substitutions between individuals, demonstrating the presence of some level of genomic heterogeneity among *Wolbachia* that infect the same *C. pipiens* field population. Overall, the absence of genetic heterogeneity within *Wolbachia* populations in a single individual confirms the presence of a dominant *Wolbachia* that is maintained under strong purifying forces of evolution.

Keywords: Wolbachia, mosquitoes, core pangenome, metapangenomics, subpopulations, single nucleotide polymorphism, single-nucleotide variants, punctual mutations

Introduction

Wolbachia is a maternally inherited intracellular bacterium widely used in biocontrol programs thanks to its ability to modulate the arthropod reproduction and to reduce the capacity to transmit pathogens [1–6] or the lifespan of pathogen host [7–13]. *Wolbachia* mainly infects the germline but also occurs in somatic tissues like fat body, hemolymph, central nervous system, for example [14]. It is mostly vertically transmitted through the female germline [15].

The endosymbiont induces multiple reproductive alterations to favor its spread by increasing the proportion of infected females (i.e. the transmitting sex) in the population: cytoplasmic incompatibility (CI) [15, 16], male killing [15–17], parthenogenesis [15, 16, 18], male feminization [15, 16]. CI is the most common reproductive manipulation and causes non-viable embryos when males infected with *Wolbachia* cross with uninfected females or when male and female are infected by incompatible *Wolbachia* variants [19–22]. In addition, transfection in *Aedes aegypti* mosquitoes with *Wolbachia* can diminish the transmission of some pathogens like Dengue, Chikungunya, or Zika [1–6]. Nevertheless, it can also enhance the transmission of others like West Nile virus [23]. Protective or reproductive phenotype disparities may be a result of species and strain-specific *Wolbachia*-host-virus interactions, which combination is also influenced by other factors like *Wolbachia* density, temperature, and host genetics, creating a system particularly difficult to disentangle.

Phylogenetic studies based on a multi-locus sequence typing (MLST) system comprised of conserved housekeeping genes [24] show that *Wolbachia* belong to at least 17 possible phylogenetic supergroups (named A-F, H-Q, and S), with the vast majority belonging to the group B-*Wolbachia* [16]. More recently, whole genome sequencing provided insights into the higher *Wolbachia* genetic diversity [16, 24–26]. In natural populations of the common house mosquito *Culex pipiens*, genotyping approaches using supplementary genes encoding proteins with ankyrin (ANK) motifs and Mobile Genetic Elements markers allowed to identify more than 100 genetically distinct *Wolbachia* variants belonging to five distinct phylogenetic groups (*w*PipI to *w*PipV) (referred to as *w*Pip strains [27–30]).

Nevertheless, most studies focusing on inter-individual variations of infections are based on a restricted set of genes that belong to the core and accessory genome, preventing comprehensive insights into the extent of homogeneity among *Wolbachia* cells within naturally infected individual hosts. *Wolbachia* are exposed to strong bottleneck effects during vertical transmission; they

Received: 12 October 2023. **Revised:** 26 April 2024. **Accepted:** 6 June 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of the International Society for Microbial Ecology.
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.

need to be transmitted to eggs, remain through embryogenesis, and finally become integrated in the founders of germ line stem cells [31] which may lead to monoclonal *Wolbachia* populations. These transmission bottlenecks are indeed shown to homogenize endosymbiotic bacteria in other systems like *Buchnera* in aphids, due to genetic drift and selection [32]. Although recently debated, the restricted niche of bacterial endosymbionts or pathogens also leads to the general assumption that only a few cells are sampled to start a new population from the millions of cells forming a within-host population [33]. Porter and Sullivan nevertheless note that *Wolbachia* could also follow a more indirect vertical transmission route by migrating from the somatic tissues to the germ line at each generation [31]. In addition, intra-individual variations of *Wolbachia* infections exist: *Wolbachia* can be horizontally transmitted [34, 35] and spread across distantly related arthropod taxa, a process that can generate co-infection of individual hosts by phylogenetically unrelated *Wolbachia* strains. Case studies include co-infection by the *w*AlbA and *w*AlbB *Wolbachia* strains in the invasive "Asian tiger" mosquito species *Aedes albopictus* [36, 37]. In addition to horizontal transfers, intra-individual structural variations of *Wolbachia* genomes have been shown. Chrostek and Teixeira for example showed intra-host variability in Octomom (a *Wolbachia* specific region including eight genes associated with density regulation) copy number between *Wolbachia* cells and within-host selection for faster replicating bacterial symbionts during the lifespan of flies [13, 38, 39]. Overall, gene and genome-level microbial population studies have been shedding light on cryptic bacterial microdiversity within single individual mosquitoes [40] or marine animals like mussels [41].

Despite the critical role of *Wolbachia* in biotechnological applications of pathogen transmission control strategies, whole genome-scale comprehensive insights into the extent of homogeneity within *Wolbachia* populations are lacking. Here, we used shotgun metagenomics to reconstruct *Wolbachia* genomes from single ovary and midgut samples obtained from adult *C. pipiens* mosquitoes collected in the South of France. We generated a pangenome to focus our analysis on the *Wolbachia* of *Culex* spp. core pangenome (genes present in single copy in reconstructed and reference genomes). We analyzed genetic variations within and between samples to investigate the putative presence of distinct *Wolbachia* populations in single individual organs and between individuals using a set of stringent filters to minimize the influence of bioinformatics artifacts.

Materials and methods

Sample collection, preparation, and sequencing

We collected and dissected individual mosquitoes, prepared four ovaries (O03, O07, O11, O12) and their corresponding midgut samples (M03, M07, M11, M12, together with two additional "orphan" samples M01, M09) for sequencing as in Reveillaud *et al.* [42] (see Supplementary Note S1 for further details).

Metagenomic assembly and binning

We performed metagenomic analyses using anvi'o v7.1 [43, 44] and the metagenomic snakemake [45] workflow from the quality filtering to the merging of profile databases generated for each organ separately using the "anvi-run-workflow" program and the "–workflow metagenomics" flag. All the parameters used to set the snakemake workflows are written in the "config.json" files available in the Data Availability section. Although the ovaries raw reads have already been analyzed in Reveillaud *et al.* [42], we herein reanalyzed both ovaries and midguts raw reads following an exactly similar protocol to generate consistent analyses and comparable results for both organs. Briefly, during each workflow, we quality-filtered the raw reads from each sample using illumina-utils [46] v2.10 and the "iu-filter-quality-minoche" anvi'o program with default parameters. We assembled quality-filtered reads into contigs using MEGAHIT [47] v1.2.9, keeping only contigs with a length > 1000 nt. We performed read recruitment analyses by mapping the quality-filtered reads from all ovary samples onto the contigs of each ovary sample with "all against all" flag in config files using Bowtie2 [48] v2.3.5.1 and repeated the same procedure for midgut samples. We then used the "anvi-gen-contigs-database" program to generate anvi'o contigs databases for each individual assembly. This program computed k-mer frequencies for each contig, soft-split contigs with lengths > 20 000 bp into smaller ones, and identified ORFs in each contig using Prodigal [49] v2.6.3. We used the "anvi-run-hmms" program to identify HMM hits searching against the default HMM sources in anvi'o (Bacteria_71, Archaea_76, and Protista_83) and the "anvi-run-ncbi-cogs" program to assign functions to genes by searching their amino acid sequences against the COG20 [50] database using blastp [51] v2.10.1. We used the "anvi-profile" program to compute the coverage per nucleotide position and statistics for each metagenome assembly using the BAM files. We merged the resulting anvi'o profiles using the "anvi-merge" program. After the metagenomic snakemake workflow, we performed an automatic genome binning from assemblies using the "anvi-cluster-contigs" anvi'o program and the CONCOCT [52] algorithm ("–driver CONCOCT" flag) with a limited number of clusters ("–clusters X" flag) by sample (Supplementary Table S1) to separate bacterial and eukaryotic reads while avoiding bacterial genome dispersion (checked with "anvi-estimate-genome-completeness" program). Finally, we manually refined the bacterial bins obtained from each sample with the interactive program "anvi-refine." In addition, we ran the "references-mode" of the metagenomic snakemake workflow to perform read recruitment of the quality-filtered reads from all samples to the refined *Wolbachia* Metagenome-Assembled Genome (MAG). We then removed low quality mappings with samtools [53] by filtering out reads with MAPQ <20, and finally performed anvi'o profiling and merged for inter-organ comparisons. Completion and redundancy of the five refined *Wolbachia* MAGs were estimated during anvi'o summary, as well as computed using the CheckM lineage workflow [54].

Pangenomics

We performed pangenomic analysis for the five *Wolbachia* MAGs obtained andthree selected *Wolbachia* reference genomes: *w*PipPel isolated from *Culex quinquefasciatus* (NCBI Accession ID NC_010981.1) [55], *w*PipMol isolated from *Culex molestus* (NCBI Accession ID NZ_CTEH00000000.1) [56], and *w*PipJHB isolated from *C. quinquefasciatus* (NCBI Accession ID NZ_ABZA00000000.1) [57]. We downloaded the fasta files of the three selected *Wolbachia* reference genomes and reformatted them using the anvi'o "anvi-script-reformat-fasta" program. We then generated a new contigs database from the reformatted fasta files using the "anvi-gen-contigs-database" program. We identified HMM hits using the "anvi-run-hmms" program and used these hmm profiles to assign functions with "anvi-run-ncbi-cogs." We created an external genome database including all these *Wolbachia* reference contigs databases. We created an internal genome database including the five *Wolbachia* MAGs contigs databases stored in the profile and contigs databases generated during the snakemake workflow with references mode. We then generated a genome storage database from both external and internal genome databases using the "anvi-gen-genomes-storage" program.

We computed the pangenome with the "anvi-pan-genome" program (using "–use-ncbi-blast" "–mcl-inflation 10" and the "genome-name" flags) and identified gene clusters for the five *Wolbachia* MAGs and three reference genomes based on amino acid sequence similarity. Only highly similar genes are added in a gene cluster during the anvi'o pangenomic workflow with almost no chance for two highly similar genes to end up in distinct gene clusters. We finally used the "anvi-display-pan" program to display the pangenome and visualize the distribution of gene clusters across genomes. From the pangenome summary, we obtained the id of gene clusters composed of genes occurring in single-copy in each genome. For convenience in the following analyses, we referred to genes belonging to these gene clusters as *Wolbachia* Single-copy Core Genes (*w*SCGs). Finally, we performed an additional sanity check on the selected *w*SCGs by confirming that their coverage was uniform over each metagenome, while the coverage of multi-copy gene clusters was variable and sparser (Supplementary Fig. S1).

Prophage WO, MLST, wsp, cidA/B putative hits

We used the available results of blastn that identified *w*PipPel genes that match WO prophage regions ("WO_in_wPip_best_hit.txt" file from https://merenlab.org/data/wolbachia-plasmid/#identifying-genes-that-correspond-to-wo-prophages detailed in Reveillaud *et al.* [42]) together with the same custom R [58] script to identify gene cluster ids corresponding to these "phage like" gene calls.

We used blastn to identify MLST (MultiLocus Sequence Typing—*gatB*, *coxA*, *hcpA*, *ftsZ*, *fbpA*) and *wsp* genes from the PubMLST [59] (Public databases for molecular typing and microbial genome diversity) in *w*PipPel. Similarly, we identified best hits using blastn for *cidA* (NCBI Accession ID from MF444963 to MF444981 for the 18 *cidA* variants) and *cidB* (NCBI Accession ID from MF444982 to MF444996 for the 14 *cidB* variants) genes from Bonneau *et al.*[60]. Finally, we identified the gene clusters corresponding to these hits in *w*PipPel.

Metapangenomics for inter-organ variability

We used a custom R script (based on https://merenlab.org/data/wolbachia-plasmid/#recovering-coverage-values-for-gene-clusters-of-the-wolbachia-pangenome-in-c-pipiens-metagenomes) to extract coverage values of metagenomes M11 and O11 mapped on MAG O11 and MAG M11 genes from the merged profile database and compute their means by gene cluster. We imported coverage values and WO prophage assignation described above to the pangenome database using the "anvi-import-misc-data" anvi'o program to build the metapangenome. We finally ran and edited the metapangenome using the "anvi-display-pangenome" program.

Single nucleotide variants, single codon variants, single amino acid variants, and single nucleotide polymorphisms

We then used the "anvi-gen-variability-profile" program to extract the tables of Single Nucleotide Variants (SNVs, "–engine NT" parameter), Single Codon Variants (SCVs, "–engine CDN"), and Single Amino Acid Variants (SAAVs, "–engine AA") from the anvi'o merged profile databases. Based on the summary from the pangenome, we added to these tables gene cluster information, including SCG/*w*SCG status and phage WO, MLST, *wsp*, and *cid*A/B putative assignation.

We quantified inter-sample variation by filtering the raw SNV tables, keeping between-sample SNVs occurring in *w*SCGs, not flagged as coverage outliers, and with a departure from the reference >0.98. These SNVs can be referred to as Single Nucleotide Polymorphisms (SNPs). We confirmed from the MAG coverage summaries that detection (or breadth of coverage) for the genes in which we found SNPs was equal to 1, to avoid partial mapping biases. We used the gene id and codon number information from the SNP tables, as well as the departure from reference >0.98 filter to obtain the associated SCV and SAAV tables from the raw tables.

We then focused our analysis on intra-sample variation by keeping only within-sample SNVs occurring in *w*SCGs, not flagged as coverage outliers (that can result from bioinformatic biases including breaks in or lack of assembly, unspecific mapping, etc.) and with entropy <0.2 and departure from consensus <0.2 (to discard those that could be due to sequencing errors and therefore considered as noise, https://merenlab.org/2013/11/04/oligotyping-best-practices/).

We visualized SNVs and SNPs through anvi'o with the « anvi-script-visualize-split-coverages » program and in Integrative Genome Viewer [61]. Finally, summary plots of the data contained in SNV and SNP tables were obtained in R. The fully reproducible workflow for this analysis is available at https://github.com/jreveillaud/Wolbachia-subpopulations.

Results

Reconstruction of *Wolbachia* MAGs in one midgut and four ovaries of *C. pipiens* individuals

Our quality filtering of raw reads sequenced from midgut and ovaries samples from individual mosquitoes resulted in 94 024 472 and 75 040 983 paired-end reads on average, respectively (Supplementary Table S2). Individual sample metagenomic assembly generated on average 166 820 contigs >1 kb recruiting between 24% and 92% of filtered reads (Supplementary Table S2). To estimate the proportion of eukaryotic reads (that we herein refer to as "contamination" in opposition to bacterial reads) in our metagenomes, we used phyloFlash [62] to annotate short reads based on the SILVA rRNA database [63] (see Supplementary Note S2 for further details). Results suggested that the vast majority of our reads (over 99% for each sample) originated from eukaryotic organisms, especially in midgut metagenomes (Supplementary Fig. S2; Supplementary Table S3).

Despite the high eukaryotic contamination rate, we reconstructed *Wolbachia* genomes from all ovary metagenomes and one of the four midgut metagenomes (M11) with 91.5% completion and 0% redundancy estimated based on Bacterial Single-Copy core Genes (BSCGs) from the collection of Campbell *et al.* [64] after manual refinement (Table 1, Supplementary Table S1). This is, to our knowledge, the first *Wolbachia* draft genome reconstructed from a *Culex* mosquito midgut. Of note, midgut metagenomes M01 and M09, which had no corresponding ovary samples, were solely used to improve binning (by providing additional differential coverage information). These samples were not further investigated as we did not reconstruct bacterial genomes from them. During the final read recruitment step, the refined MAGs recruited between 0.83% and 3.48% of reads in the metagenomes they were respectively reconstructed from.

Comparison of *Wolbachia* MAGs between organs of the same individual

As we reconstructed for the first time a *Wolbachia* MAG from a midgut metagenome, we investigated the putative occurrence of organ-specific gene clusters at the individual level. We first observed 19 gene clusters that seemed to be unique to *Wolbachia*

Table 1. Refined *Wolbachia* MAGs estimates including completion and redundancy rates, number of contigs, total number of nucleotides (length), and GC content.

Wolbachia MAG	Completion (%) (BSCGs)	Redundancy (%) (BSCGs)	Completion (%) (CheckM)	Redundancy (%) (CheckM)	Number of contigs	Length (bp)	GC content (%)
M11	91.55	0	100	0.09	138	1 331 260	34.2
O11	91.55	0	100	0.09	119	1 290 070	34.2
O03	91.55	0	99.15	0.09	73	1 164 954	33.8
O07	91.55	0	100	0.09	143	1 340 038	34.4
O12	91.55	0	99.15	0.09	75	1 181 440	33.9

The MAGs showed a high completion >90% and no redundancy based on the use of Bacterial Single-Copy core Genes (BSCGs) from the collection of Campbell et al. [64] and the use of CheckM [54].

MAG M11 and four gene clusters possibly unique to *Wolbachia* MAG O11 (Supplementary Fig. S3; Supplementary Table S4). Nevertheless, in our metapangenomic analysis, the mapping of quality filtered reads onto the two *Wolbachia* MAGs showed that those gene clusters had coverage in all samples and thus were not unique to one specific *Wolbachia* MAG (Supplementary Fig. S3). The absence of some genes in our *Wolbachia* MAGs could be explained by assembly breaks and/or the exclusion of contigs with length < 1000 bp. We therefore did not observe clear evidence of organ-specific *Wolbachia* populations (Supplementary Table S4; Supplementary Fig. S3).

Single-copy core genes in *Wolbachia* MAGs and reference genomes

Furthermore, we compared gene content between our five newly reconstructed *Wolbachia* MAGs and three *Wolbachia* reference genomes (*w*PipPel, *w*PipMol, and *w*PipJHB) through pangenomic analyses. Overall, we identified 1205 gene clusters (Supplementary Table S5), among which 890 were single-copy gene clusters, i.e. composed of a single gene sequence for each *Wolbachia* MAG and reference genome. The sequences belonging to these single-copy gene clusters were referred to as *Wolbachia* Single-copy Core Genes (*w*SCGs) within each MAG (Supplementary Table S6) as discussed in the Material and Methods section. To note, we identified *w*SCGs corresponding to MLST and *wsp* genes but not to any *cid* gene (Supplementary Table S6).

Wolbachia population genetics between individual mosquitoes (inter-sample variability)

We looked for SNPs, i.e. variable positions with 100% divergence from the reference sequence, to investigate the possible presence of fixed mutations between individuals within *w*SCGs. After recruiting reads from all metagenomes to the five reconstructed MAGs, we filtered our mapping results to only keep reads with a mapping quality over 20. This removed between 1.13% and 2.92% of recruited reads, losing some information but increasing the robustness of our analysis (Supplementary Fig. S4). We then selected point mutations in *w*SCGs with departure from reference over 0.98, always making sure that they were not identified as coverage outliers. Finally, we checked that detection (or breadth of coverage) was equal to 1 on the considered genes to avoid errors due to partial recruitment.

We identified SNPs for all inter-individual comparisons, except when mapping metagenome M11 to *Wolbachia* MAG reconstructed from O11, and vice-versa (Supplementary Tables S7 and S8). In total, SNPs were identified in 23 gene clusters, with 22 variable positions in MAGs M11, O03, O07, O11 and 23 in MAG O12 (Fig. 1A; Supplementary Table S9). All SNPs gave rise to SCVs (Supplementary Table S10) and a number of them resulted in SAAVs (Supplementary Table S11), with a mean SAAV to SCV ratio between 0.75 and 0.82 (Supplementary Table S12).

Wolbachia population genetics within individual mosquitoes (intra-sample variability)

In contrast to the small number of inter-individual SNPs across genomes we described above, the raw SNVs from metagenomic read recruitment results suggested a remarkable number of intra-individual variants. However, a higher level of scrutiny revealed that these variants in the core pangenome could be attributed to bioinformatic artifacts (See Supplementary Note S3, Supplementary Figs S8–14 and Table S13 for further details of the analysis and visualizations).

Discussion

We reconstructed *Wolbachia* MAGs from both the ovaries and midgut of one *C. pipiens* individual for the first time, as well as from the ovaries of three additional *C. pipiens* specimen using metagenomic approaches. Our metapangenomic analyses indicated that ovaries and midgut from a single mosquito share similar *Wolbachia* gene content, suggesting the uniform segregation and the lack of strain selection across organs. In addition, variability analyses at the inter-sample level showed the existence of synonymous and non-synonymous SNPs, with different occurrence patterns across individuals, suggesting fixed punctual mutations and multiple *Wolbachia* populations. However, a detailed SNV investigation within *w*SCG at the intra-sample level showed the absence of punctual mutations.

Globally, *Wolbachia* is manipulated with the idea that it is monoclonal in transfection and naturally infected mosquito specimen. *Wolbachia* is predominantly extracted from egg cytoplasm of an infected species before being transferred to a recipient one [65] during transfection. Although a relative stability of *Wolbachia* genomes has been observed following the artificial transfer of the bacterium between host species for several years, higher mutation rates were recently shown in *A. aegypti* cell lines, suggesting that different population dynamics can occur following distinct selective pressures within specific environments [66]. Similarly, the action of selective sweep has been documented on *Wolbachia* genomes from *Drosophila melanogaster* [67]. The absence of genetic heterogeneity shown here in the *Wolbachia* core pangenome within single naturally-infected mosquito organs and specimen is congruent with evolutionary processes acting against mutations within samples, including reproductive bottleneck and a strong purifying selection. In addition, we did not detect different gene content nor any SNPs in *Wolbachia* from different organs of the same individual, highlighting the uniformity of *Wolbachia* at the mosquito level. Our data agree with a single *Wolbachia* population

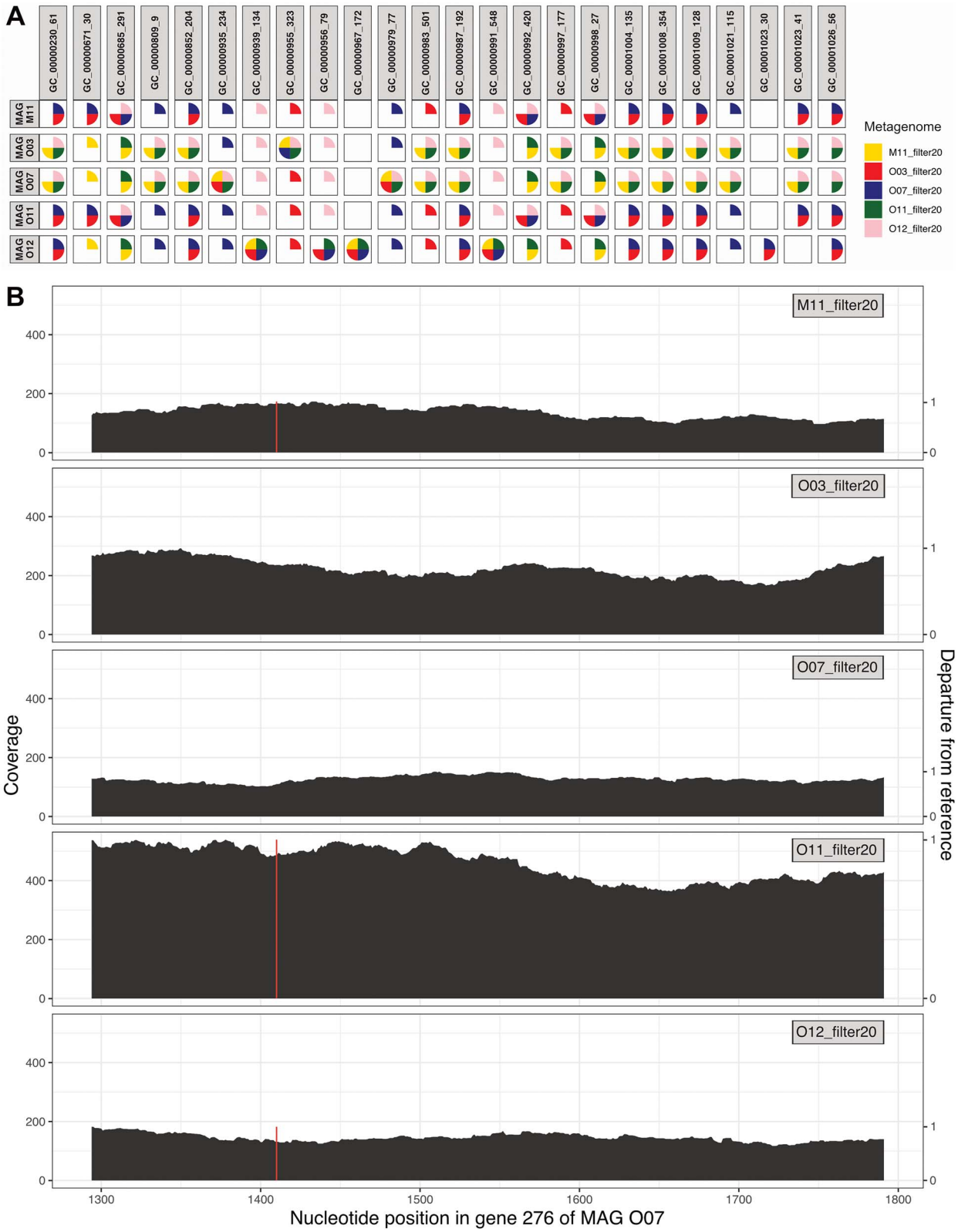

Figure 1. SNPs identification and visualization; (A) representation of the variable positions (vertical name corresponding to gene cluster id + codon id in which it occurs) in the five *Wolbachia* MAGs and the metagenomes in which a SNP is identified (colored pie chart); (B) visualization of gene 276 (gene cluster GC_00001009) from contig 124 766 in MAG O07; a SNP is identified in metagenomes M11, O11, and O12, in first position in codon 128 (represented by a contrasting bar); this gene was annotated as « Holliday junction resolvasome RuvABC endonuclease subunit RuvC » by COG20; additional visualizations of SNPs are available in Supplementary Figs S5–S7.

that is transferred from the mother to the offspring [14, 68] and then from the germ line to the somatic tissue.

The observation of SNPs, differentially co-occurring across individuals, and in some cases non-synonymous, nevertheless question the emergence and evolution of variants. As of now, the evolutionary processes giving rise to these fixed mutations remain unknown. Theoretically, non-neutral processes could drive the emergence of distinct variants conferring evolutionary

advantages to their host, such as protection against pathogens in tripartite *Wolbachia*-host-pathogen interactions [69]. However, it could just as likely result from drift and fixation in the progeny through a random transmission event. These processes would be studied most efficiently by monitoring *Wolbachia* evolution in the progeny of an iso-female line over a long period of time.

Despite the first striking identification of SNVs in *w*SCG genes within samples (Supplementary Table S7), a close examination of SNVs and coverage variations highlighted cryptic and hidden bioinformatic bias, most likely due to the fragmented nature of *Wolbachia* MAGs. Indeed, although we focused our analysis on SNVs occurring only within *w*SCG (that showed a single copy gene signature using a combined pangenomic and metapangenomic approach), an in-depth investigation revealed the occurrence of SNVs significantly correlated with a subtle increase in coverage. Blast outputs confirmed these bioinformatic artefacts, suggesting these data were due to (i) hidden conserved domains within target genes, as well as (ii) genes that were not reconstructed in fragmented *Wolbachia* genomes despite high completion values (91.5% to 100% depending on anvi'o or CheckM estimates). Indeed, a high number of transposable elements render the obtention of circular *Wolbachia* genomes particularly challenging [70]. In addition, ANK repeat domain encoding genes, particularly numerous in *Wolbachia* genomes (23 in *w*Mel, 60 in *w*Pip strain [71]), could impede assembly and consequently favor non-specific read recruitment. Similar patterns of unspecific read recruitment could be observed for other intracellular bacteria including pathogens like *Ehrlichia*, which shows a high number of tandem repeats [72].

Overall, making good use of *Wolbachia* requires information on the genetic variation of the host, the pathogen, and the endosymbiont at fine scale, as distinct variants can alter pathogen virulence as well as the efficiency of the protective or reproductive phenotype. *Wolbachia* is widely used in antivectorial programs worldwide to fight diseases, and knowledge of bacterial diversity within and between single individuals is critical. Here our analysis focused on the core pangenome of *Wolbachia* due to the type of data we were working with (short read). It would be beneficial to extend it to the whole genome using other techniques such as long-read sequencing that could yield less fragmented genomes and allow studying structural variations at the individual scale.

Acknowledgements

We are thankful to Valentin Loux, Cédric Midoux, Véronique Martin for their support on the MIGALE high-computer during the data analysis, Corinne Cruaud from the Genoscope for her assistance with sequencing, as well as Ewa Chrostek and Wolfgang Miller for helpful discussion about *Wolbachia* variants. We also thank EID Mediterranée and Ignace Rakotoarivony for mosquito sampling and dissections, respectively.

Author contributions

Blandine Trouche and Hans Schrieke analyzed the data, generated the tables and figures, and wrote the manuscript. Olivier Duron and A. Murat Eren analyzed the data and wrote the manuscript. Julie Reveillaud coordinated the study, analyzed the data, and wrote the manuscript. All authors have read, contributed, and approved the final version of the manuscript.

Supplementary material

Supplementary material is available at *ISME Communications* online.

Conflicts of interest

The authors declare no competing interests.

Funding

This work was supported by the ERC RosaLind Starting Grant "948135" to J.R., and the INRAe Animal Health Department and Occitany Region to H.S. and J.R. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability

The raw sequencing data for the shotgun metagenomes of midgut and ovary samples are available in the European Nucleotide Archive via accession number PRJEB56379 and PRJEB26028, respectively. In addition, we made available the merged anvi'o profiles for the midgut and ovary metagenomes (https://doi.org/10.5281/zenodo.7183277), the FASTA files for the five *Wolbachia* MAGs (https://doi.org/10.5281/zenodo.7183303), the anvi'o merged profile databases for the *Wolbachia* MAGs used for the SNVs and metapangenomic analyses (https://doi.org/10.5281/zenodo.7183324).

A reproducible bioinformatics workflow including scripts used for all computational analyses is available at the URL https://github.com/jreveillaud/Wolbachia-subpopulations (https://doi.org/10.5281/zenodo.11059970).

References

1. Moreira LA, Iturbe-Ormaetxe I, Jeffery JA *et al.* A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, chikungunya, and plasmodium. *Cell* 2009;**139**:1268–78. https://doi.org/10.1016/j.cell.2009.11.042
2. van den Hurk AF, Hall-Mendelin S, Pyke AT *et al.* Impact of *Wolbachia* on infection with chikungunya and yellow fever viruses in the mosquito vector *Aedes aegypti*. *PLoS Negl Trop Dis* 2012;**6**:e1892. https://doi.org/10.1371/journal.pntd.0001892
3. Caragata EP, Dutra HLC, Moreira LA. Inhibition of Zika virus by *Wolbachia* in *Aedes aegypti*. *Microb Cell* 2016;**3**:293–5. https://doi.org/10.15698/mic2016.07.513
4. Dutra HLC, Rocha MN, Dias FBS *et al.* *Wolbachia* blocks currently circulating Zika virus isolates in Brazilian *Aedes aegypti* mosquitoes. *Cell Host Microbe* 2016;**19**:771–4. https://doi.org/10.1016/j.chom.2016.04.021
5. Aliota MT, Peinado SA, Velez ID *et al.* The *w*Mel strain of *Wolbachia* reduces transmission of Zika virus by *Aedes aegypti*. *Sci Rep* 2016;**6**:28792. https://doi.org/10.1038/srep28792
6. Pereira TN, Rocha MN, Sucupira PHF *et al.* *Wolbachia* significantly impacts the vector competence of *Aedes aegypti* for Mayaro virus. *Sci Rep* 2018;**8**:6889. https://doi.org/10.1038/s41598-018-25236-8
7. Flores HA, O'Neill SL. Controlling vector-borne diseases by releasing modified mosquitoes. *Nat Rev Microbiol* 2018;**16**:508–18. https://doi.org/10.1038/s41579-018-0025-0
8. Hoffmann AA, Montgomery BL, Popovici J *et al.* Successful establishment of *Wolbachia* in Aedes populations to suppress dengue transmission. *Nature* 2011;**476**:454–7. https://doi.org/10.1038/nature10356
9. Huang Y-JS, Higgs S, Vanlandingham DL. Biological control strategies for mosquito vectors of arboviruses. *Insects* 2017;**8**:21. https://doi.org/10.3390/insects8010021
10. Benelli G, Jeffries C, Walker T. Biological control of mosquito vectors: past, present, and future. *Insects* 2016;**7**:52. https://doi.org/10.3390/insects7040052

11. Zheng X, Zhang D, Li Y *et al.* Incompatible and sterile insect techniques combined eliminate mosquitoes. *Nature* 2019;**572**:56–61. https://doi.org/10.1038/s41586-019-1407-9
12. McMeniman CJ, Lane RV, Cass BN *et al.* Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti*. *Science* 2009;**323**:141–4. https://doi.org/10.1126/science.1165326
13. Chrostek E, Marialva MSP, Esteves SS *et al.* *Wolbachia* variants induce differential protection to viruses in *Drosophila melanogaster*: a phenotypic and phylogenomic analysis. *PLoS Genet* 2013;**9**:e1003896. https://doi.org/10.1371/journal.pgen.1003896
14. Pietri JE, DeBruhl H, Sullivan W. The rich somatic life of *Wolbachia*. *Microbiologyopen* 2016;**5**:923–36. https://doi.org/10.1002/mbo3.390
15. Werren JH, Baldo L, Clark ME. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* 2008;**6**:741–51. https://doi.org/10.1038/nrmicro1969
16. Kaur R, Shropshire JD, Cross KL *et al.* Living in the endosymbiotic world of *Wolbachia*: a centennial review. *Cell Host Microbe* 2021;**29**:879–93. https://doi.org/10.1016/j.chom.2021.03.006
17. Perlmutter JI, Bordenstein SR, Unckless RL *et al.* The phage gene wmk is a candidate for male killing by a bacterial endosymbiont. *PLoS Pathog* 2019;**15**:e1007936. https://doi.org/10.1371/journal.ppat.1007936
18. Stouthamer R, Breeuwert JA, Luck RF *et al.* Molecular identification of microorganisms associated with parthenogenesis. *Nature* 1993;**361**:66–8. https://doi.org/10.1038/361066a0
19. Shropshire JD, Leigh B, Bordenstein SR. Symbiont-mediated cytoplasmic incompatibility: what have we learned in 50 years? *eLife* 2020;**9**:e61989. https://doi.org/10.7554/eLife.61989
20. LePage DP, Metcalf JA, Bordenstein SR *et al.* Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. *Nature* 2017;**543**:243–7. https://doi.org/10.1038/nature21391
21. Beckmann JF, Bonneau M, Chen H *et al.* The toxin–antidote model of cytoplasmic incompatibility: genetics and evolutionary implications. *Trends Genet* 2019;**35**:175–85. https://doi.org/10.1016/j.tig.2018.12.004
22. Sicard M, Bonneau M, Weill M. *Wolbachia* prevalence, diversity, and ability to induce cytoplasmic incompatibility in mosquitoes. *Curr Opin Insect Sci* 2019;**34**:12–20. https://doi.org/10.1016/j.cois.2019.02.005
23. Dodson BL, Hughes GL, Paul O *et al.* *Wolbachia* enhances West Nile virus (WNV) infection in the mosquito *Culex tarsalis*. *PLoS Negl Trop Dis* 2014;**8**:e2965. https://doi.org/10.1371/journal.pntd.0002965
24. Baldo L, Dunning Hotopp JC, Jolley KA *et al.* Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Appl Environ Microbiol* 2006;**72**:7098–110. https://doi.org/10.1128/AEM.00731-06
25. Kaur R, Siozios S, Miller W *et al.* Insertion sequence polymorphism and genomic rearrangements uncover hidden *Wolbachia* diversity in *Drosophila suzukii* and *D. subpulchrella*. *Sci Rep* 2017;**7**:14815. https://doi.org/10.1038/s41598-017-13808-z
26. Scholz M, Albanese D, Tuohy K *et al.* Large scale genome reconstructions illuminate *Wolbachia* evolution. *Nat Commun* 2020;**11**:5235. https://doi.org/10.1038/s41467-020-19016-0
27. Atyame CM, Delsuc F, Pasteur N *et al.* Diversification of *Wolbachia* endosymbiont in the *Culex pipiens* mosquito. *Mol Biol Evol* 2011;**28**:2761–72. https://doi.org/10.1093/molbev/msr083
28. Duron O, Fort P, Weill M. Hypervariable prophage WO sequences describe an unexpected high number of *Wolbachia* variants in the mosquito *Culex pipiens*. *Proc R Soc B Biol Sci* 2006;**273**:495–502. https://doi.org/10.1098/rspb.2005.3336
29. Duron O, Boureux A, Echaubard P *et al.* Variability and expression of ankyrin domain genes in *Wolbachia* variants infecting the mosquito *Culex pipiens*. *J Bacteriol* 2007;**189**:4442–8. https://doi.org/10.1128/JB.00142-07
30. Duron O, Raymond M, Weill M. Many compatible *Wolbachia* strains coexist within natural populations of *Culex pipiens* mosquito. *Heredity* 2011;**106**:986–93. https://doi.org/10.1038/hdy.2010.146
31. Porter J, Sullivan W. The cellular lives of *Wolbachia*. *Nat Rev Microbiol* 2023;**21**:750–66. https://doi.org/10.1038/s41579-023-00918-x
32. Mira A, Moran NA. Estimating population size and transmission bottlenecks in maternally transmitted Endosymbiotic bacteria. *Microb Ecol* 2002;**44**:137–43. https://doi.org/10.1007/s00248-002-0012-9
33. Stritt C, Gagneux S. How do monomorphic bacteria evolve? The *mycobacterium tuberculosis* complex and the awkward population genetics of extreme clonality. *Eco EvoRxiv* 2023;**4829**:ver. 3, peer-reviewed and recommended by Peer Community in Evolutionary Biology. https://doi.org/10.32942/X2GW2P
34. Brown AN, Lloyd VK. Evidence for horizontal transfer of *Wolbachia* by a Drosophila mite. *Exp Appl Acarol* 2015;**66**:301–11. https://doi.org/10.1007/s10493-015-9918-z
35. Raychoudhury R, Baldo L, Oliveira DCSG *et al.* Modes of Acquisition of *Wolbachia*: horizontal transfer, hybrid introgression, and codivergence in the Nasonia species complex. *Evolution* 2009;**63**:165–83. https://doi.org/10.1111/j.1558-5646.2008.00533.x
36. Dutton TJ, Sinkins SP. Strain-specific quantification of *Wolbachia* density in *Aedes albopictus* and effects of larval rearing conditions. *Insect Mol Biol* 2004;**13**:317–22. https://doi.org/10.1111/j.0962-1075.2004.00490.x
37. Calvitti M, Marini F, Desiderio A *et al.* *Wolbachia* density and cytoplasmic incompatibility in *Aedes albopictus*: concerns with using artificial *Wolbachia* infection as a vector suppression tool. *PLoS One* 2015;**10**:e0121813. https://doi.org/10.1371/journal.pone.0121813
38. Chrostek E, Teixeira L. Within host selection for faster replicating bacterial symbionts. *PLoS One* 2018;**13**:e0191530. https://doi.org/10.1371/journal.pone.0191530
39. Chrostek E, Teixeira L. Mutualism breakdown by amplification of *Wolbachia* genes. *PLoS Biol* 2015;**13**:e1002065. https://doi.org/10.1371/journal.pbio.1002065
40. Schrieke H, Maignien L, Constancias F *et al.* The mosquito microbiome includes habitat-specific but rare symbionts. *Comput Struct Biotechnol J* 2022;**20**:410–20. https://doi.org/10.1016/j.csbj.2021.12.019
41. Ikuta T, Takaki Y, Nagai Y *et al.* Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population. *ISME J* 2016;**10**:990–1001. https://doi.org/10.1038/ismej.2015.176
42. Reveillaud J, Bordenstein SR, Cruaud C *et al.* The *Wolbachia* mobilome in *Culex pipiens* includes a putative plasmid. *Nat Commun* 2019;**10**:1051. https://doi.org/10.1038/s41467-019-08973-w
43. Eren AM, Esen ÖC, Quince C *et al.* Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 2015;**3**:e1319. https://doi.org/10.7717/peerj.1319
44. Eren AM, Kiefl E, Shaiber A *et al.* Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* 2021;**6**:3–6. https://doi.org/10.1038/s41564-020-00834-3
45. Koster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* 2012;**28**:2520–2. https://doi.org/10.1093/bioinformatics/bts480

46. Eren AM, Vineis JH, Morrison HG *et al.* A filtering method to generate high quality short reads using illumina paired-end technology. *PLoS One* 2013;**8**:e66643. https://doi.org/10.1371/journal.pone.0066643
47. Li D, Liu C-M, Luo R *et al.* MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015;**31**:1674–6. https://doi.org/10.1093/bioinformatics/btv033
48. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;**9**:357–9. https://doi.org/10.1038/nmeth.1923
49. Hyatt D, Chen G-L, LoCascio PF *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;**11**:119. https://doi.org/10.1186/1471-2105-11-119
50. Galperin MY, Wolf YI, Makarova KS *et al.* COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res* 2021;**49**:D274–81. https://doi.org/10.1093/nar/gkaa1018
51. Altschul SF, Gish W, Miller W *et al.* Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10. https://doi.org/10.1016/S0022-2836(05)80360-2
52. Alneberg J, Bjarnason BS, de Bruijn I *et al.* Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014;**11**:1144–6. https://doi.org/10.1038/nmeth.3103
53. Danecek P, Bonfield JK, Liddle J *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* 2021;**10**:giab008. https://doi.org/10.1093/gigascience/giab008
54. Parks DH, Imelfort M, Skennerton CT *et al.* CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;**25**:1043–55. https://doi.org/10.1101/gr.186072.114
55. Klasson L, Walker T, Sebaihia M *et al.* Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. *Mol Biol Evol* 2008;**25**:1877–87. https://doi.org/10.1093/molbev/msn133
56. Pinto SB, Stainton K, Harris S *et al.* Transcriptional regulation of *Culex pipiens* mosquitoes by *Wolbachia* influences cytoplasmic incompatibility. *PLoS Pathog* 2013;**9**:e1003647. https://doi.org/10.1371/journal.ppat.1003647
57. Salzberg SL, Puiu D, Sommer DD *et al.* Genome sequence of the *Wolbachia* endosymbiont of *Culex quinquefasciatus* JHB. *J Bacteriol* 2009;**191**:1725. https://doi.org/10.1128/JB.01731-08
58. R Core Team. *R: A language and environment for statistical computing*. R foundation for Statistical Computing, Vienna, Austria. 2023. https://www.r-project.org/
59. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 2018;**3**:124. https://doi.org/10.12688/wellcomeopenres.14826.1
60. Bonneau M, Atyame C, Beji M *et al.* *Culex pipiens* crossing type diversity is governed by an amplified and polymorphic operon of *Wolbachia*. *Nat Commun* 2018;**9**:319. https://doi.org/10.1038/s41467-017-02749-w
61. Robinson JT, Thorvaldsdóttir H, Winckler W *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011;**29**:24–6. https://doi.org/10.1038/nbt.1754
62. Gruber-Vodicka HR, Seah BKB, Pruesse E *et al.* phyloFlash: rapid small-subunit rRNA profiling and targeted assembly from metagenomes. *mSystems* 2020;**5**:e00920-20/msystems/5/5/mSys.00920-20.atom. https://doi.org/10.1128/mSystems.00920-20
63. Quast C, Pruesse E, Yilmaz P *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;**41**:D590–6. https://doi.org/10.1093/nar/gks1219
64. Campbell J, O'Donoghue P, Schwientek P *et al.* UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Proc Natl Acad Sci U S A* 2013;**110**:5540–5. https://doi.org/10.1073/pnas.1303090110
65. Hughes GL, Rasgon JL. Transinfection: a method to investigate *Wolbachia*–host interactions and control arthropod-borne disease. *Insect Mol Biol* 2014;**23**:141–51. https://doi.org/10.1111/imb.12066
66. Martinez J, Sinkins SP. Rapid evolution of *Wolbachia* genomes in mosquito cell culture. 2023; bioRxiv. https://doi.org/10.1101/2023.09.20.558649
67. Richardson MF, Weinert LA, Welch JJ *et al.* Population genomics of the Wolbachia endosymbiont in Drosophila melanogaster. *PLoS Genet* 2012;**8**:e1003129. https://doi.org/10.1371/journal.pgen.1003129
68. Veneti Z, Clark ME, Karr TL *et al.* Heads or tails: host-parasite interactions in the *Drosophila-Wolbachia* system. *Appl Environ Microbiol* 2004;**70**:5366–72. https://doi.org/10.1128/AEM.70.9.5366-5372.2004
69. Vavre F, Charlat S. Making (good) use of *Wolbachia*: what the models say. *Curr Opin Microbiol* 2012;**15**:263–8. https://doi.org/10.1016/j.mib.2012.03.005
70. Cerveau N, Leclercq S, Leroy E *et al.* Short- and long-term evolutionary dynamics of bacterial insertion sequences: insights from *Wolbachia* endosymbionts. *Genome Biol Evol* 2011;**3**:1175–86. https://doi.org/10.1093/gbe/evr096
71. Siozios S, Ioannidis P, Klasson L *et al.* The diversity and evolution of *Wolbachia* ankyrin repeat domain genes. *PLoS One* 2013;**8**:e55390. https://doi.org/10.1371/journal.pone.0055390
72. Frutos R, Viari A, Ferraz C *et al.* Comparative genomic analysis of three strains of *Ehrlichia ruminantium* reveals an active process of genome size plasticity. *J Bacteriol* 2006;**188**:2533–42. https://doi.org/10.1128/JB.188.7.2533-2542.2006