The Wolbachia mobilome in Culex pipiens includes a putative plasmid

Julie Reveillaud1, Sarah R. Bordenstein2, Corinne Cruaud3, Alon Shaiber4,5, Özcan C. Esen5, Mylène Weill6, Patrick Makoundou6, Karen Lolans5, Andrea R. Watson5, Ignace Rakotoarivony1, Seth R. Bordenstein2,7,8 & A. Murat Eren4,5,9

Wolbachia is a genus of obligate intracellular bacteria found in nematodes and arthropods worldwide, including insect vectors that transmit dengue, West Nile, and Zika viruses. Wolbachia’s unique ability to alter host reproductive behavior through its temperate bacteriophage WO has enabled the development of new vector control strategies. However, our understanding of Wolbachia’s mobilome beyond its bacteriophages is incomplete. Here, we reconstruct near-complete Wolbachia genomes from individual ovary metagenomes of four wild Culex pipiens mosquitoes captured in France. In addition to viral genes missing from the Wolbachia reference genome, we identify a putative plasmid (pWCP), consisting of a 9.23-kbp circular element with 14 genes. We validate its presence in additional Culex pipiens mosquitoes using PCR, long-read sequencing, and screening of existing metagenomes. The discovery of this previously unrecognized extrachromosomal element opens additional possibilities for genetic manipulation of Wolbachia.
Mosquitoes are major vectors of disease-causing pathogens worldwide including viruses such as dengue, West Nile, Chikungunya, Zika, and yellow fever. In the absence of effective vaccines and the off-target effects of insecticides used to control mosquitoes, novel vector biocontrol efforts are the focus of intense study. Over the past decade, the widely distributed endosymbiotic alphaproteobacteria Wolbachia has gained attention as a promising mosquito-control strategy. The basic reasons for its heightened attention is that Wolbachia can cause reproductive parasitism whereby the intracellular bacteria in the reproductive tissues can alter sexual reproduction to enhance its maternal spread through host populations at the expense of host fitness. Second and notably, native Wolbachia reduce RNA virus replication in the fruit fly Drosophila, and Wolbachia transfections into mosquito species (i.e., Aedes aegypti, Ae. albopictus, Ae. polynesiensis, and Anopheles stephensi) similarly result in mosquito lines refractory to various types of pathogenic RNA viruses. The combination of a selfish drive system and pathogen blocking by Wolbachia has led to successful pilot trials for suppression of mosquito population size and replacement of infected mosquito populations so they can no longer transmit pathogens.

Wolbachia are transovarially transmitted from the mother to offspring. In some arthropods, including the naturally infected vector species complex Culex pipiens, Wolbachia ‘modify’ sperm in testes, leading to embryonic lethality if the infected male mates with either an uninfected female or a female harbouring an incompatible Wolbachia strain. When both male and female are infected with the same Wolbachia, the modification is ‘rescued’, and compatibility is restored. This reproductive alteration, termed ‘cytoplasmic incompatibility’ (CI), comprises the most common form of Wolbachia-induced parasitism and, as studied in C. pipiens, can lead to highly diverse unidirectional and bidirectional incompatibility phenotypes. CI is currently used in vector control studies for population replacement by Wolbachia-infected strains that block arbovirus transmission, or population suppression that reduces the number of mosquito vectors. Notably, the genes responsible for CI occur in the eukaryotic association module of Wolbachia’s temperate phage WO, which laterally transfers between Wolbachia infections and evolves rapidly.

Overall, these findings emphasize the importance of further investigating mobile genetic elements in Wolbachia.

While recent studies shed light on the role of phage WO in Wolbachia genome evolution and CI, other extrachromosomal elements, such as plasmids, have not been detected in the symbiont. Notably, over half of the species in the closely related Rickettsia genus have plasmids that play roles in DNA replication, partitioning, mobilization, and conjugation and offer a potential tool for genetic manipulation of diverse members of Rickettsia. Similar genetic manipulation strategies for Wolbachia are conceivable; however, previous efforts to search for such extrachromosomal mobile genetic elements have not been successful. The lack of isolates limit direct insights into Wolbachia genomics, and most metagenomic approaches thus far rely on pooled individuals grown in the laboratory environment due to low infection densities. Since these limitations can conceal naturally occurring genomic diversity among Wolbachia populations, highly resolved analyses of individual mosquitoes may reveal additional insights into the Wolbachia ‘mobilome’, the pool of all mobile genetic elements associated with Wolbachia populations.

Here we sequence ovary samples from four wild-caught C. pipiens individuals captured in Southern France from a single trap. Using genome-resolved metagenomic and pan genomic analysis strategies, we were able to reconstruct and compare near-complete Wolbachia genomes from each individual. Besides a diverse set of virus-associated genes that were missing or absent in the reference Wolbachia genome wPip Pel, our data reveal the first lines of evidence for an extrachromosomal circular element with genetic and functional hallmarks of a plasmid that we tentatively name pWCP.

Results

A Wolbachia metagenome-assembled genome (MAG) is recovered in each sample. Shotgun sequencing of DNA recovered from ovary samples of four C. pipiens individuals (O03, O07, O11, O12) resulted in 65–78 million paired-end sequences after quality filtering. Metagenomic assembly of each sample individually yielded 147K–183K contiguous DNA segments (contigs) >1 kbp, which recruited 48.1–72.9% of the raw sequencing reads. The relatively high fraction of unmapped reads were likely due to challenges associated with the assembly of environmental metagenomes, especially in the presence of eukaryotic host genomic DNA. Supplementary Table 1 reports statistics for the raw number of reads and assembly results for each sample. We employed a metagenomic binning strategy that uses sequence composition signatures and differential coverage statistics of contigs across samples. For each ovary sample, we were able to reconstruct a highly complete single bacterial MAG that resolved to Wolbachia (Table 1).

Wolbachia MAGs include highly covered non-phage contigs. The relatively low number of single-nucleotide variants (0.01–0.05%, Supplementary Table 2) suggested that each Wolbachia MAG represented a nearly monoclonal population of bacterial cells in the ovary metagenomes. In addition, the high average nucleotide identity across our MAGs and the 1482 kbp reference genome for Wolbachia, wPip Pel (99.1–99.98%, Supplementary Table 3) suggested a high degree of conservation between the endosymbionts of different individuals. Our metagenomic read recruitment analyses using Wolbachia MAGs revealed consistent coverage statistics averaging 168×–491× for contigs that were enriched with bacterial genes. However, a subset of contigs in each individual displayed approximately a five-fold increase in coverage (Supplementary Table 4). Because Wolbachia harbour prophage WO that can enter the lytic cycle

| MAG | Percent completion (PC) | Percent redundancy (PR) | Number of contigs (N) | Number of genes (n) | Length (total number of nucleotides) | GC content (%) |
|-----|------------------------|------------------------|----------------------|---------------------|-------------------------------------|---------------|
| O03 | 94.24                  | 0.72                   | 93                   | 1091                | 1,213,072                           | 33.83         |
| O07 | 94.24                  | 0.72                   | 127                  | 1181                | 1,271,313                           | 33.78         |
| O11 | 94.24                  | 0                       | 99                   | 1085                | 1,208,099                           | 33.84         |
| O12 | 94.24                  | 0                       | 99                   | 1113                | 1,237,800                           | 33.95         |

Estimates include completion and redundancy estimates, number of contigs (N), number of genes (n), total number of nucleotides and percentage of GC. More than 90% completion and <10% redundancy based on the single occurrence of 139 single-copy genes (SCG) identified from the collection by authors in ref. 17 suggest high completion of the bins.
and form phage particles\(^{27,44,45}\), we postulated that some of the contigs could be phage associated, which would explain coverage inconsistencies. We used the five prophage regions identified in the wPip Pel reference genome\(^{27,46}\) to identify contigs enriched with genes of phage origin. Contigs classified and validated through homology searches against phage WO matched to contigs that were highly covered in our MAGs, confirming that most shifts in coverage could be attributed to phages of Wolbachia. However, surprisingly, five contigs in our MAGs (Contig O12_A, Contig O11_A, Contig O07_A, Contig O07_A’, and Contig O03_A, Supplementary Data 1) showed no homology to prophage WO despite their remarkable coverage that ranged between 720× and 2176×. Interestingly, the 5’ and 3’ ends of these contigs showed homology to the non-coding flanking regions of wPip’s ISWpi12 transposable element (TE; WP0440, WP1209, and WP1347) of the IS110-family\(^{46}\). Given their (1) high coverage in metagenomes, (2) lack of homology to prophage WO, and (3) putative association with IS110 TE, we hypothesized that these contigs could represent extrachromosomal elements.

**pWCP: a Wolbachia-associated putative plasmid.** Based on homology between the ends of these five contigs and the flanking regions of IS110 TE, we predicted that the missing region was the latter element. We artificially circularized Contig O11_A (8037 bp) and inserted an IS110 TE (1386 bp) based on the overlapping 5’ and 3’ ends (Fig. 1). Metagenomic read recruitment onto this artificially circularized contig, which we tentatively name ‘pWCP’ (for plasmid of Wolbachia endosymbiont in C. pipiens), showed consistent coverage over its entire length except a clear two-fold coverage increase in a region that matched to the IS110 TE in all four C. pipiens individuals in our study (Supplementary Figure 1). These data suggested the IS110 TE are located in the extrachromosomal pWCP while some others could be integrated in the bacterial chromosome. The read recruitment from three C. pipiens egg metagenomes generated in a previous study\(^{32}\) confirmed the near identical presence of pWCP in all three, including the increase in coverage matching to the IS110 TE (Supplementary Figure 1).

To validate the circular and extrachromosomal nature of pWCP independently of short-read recruitment- and assembly-based strategies, we generated long reads from additional C. pipiens complex samples using a MinION sequencer. Since MinION sequencing occurs with no polymerase chain reaction (PCR) or downstream assembly steps, we hypothesized that long reads that match to pWCP should never be (1) flanked by genomic regions matching to the Wolbachia chromosome and (2) longer than the pWCP itself. Our MinION sequencing analysis resulted in 14,808 high-quality sequences that were >5000 nucleotides. While a significant fraction of these reads were eukaryotic contamination and the lambda phage DNA that we used to pad our low-biomass samples (see Methods), a local BLAST search of artificially circularized pWCP sequence against long reads revealed 19 that aligned to pWCP with an e value of <1e−20 (Supplementary Data 2). Thirteen of these 19 reads covered >50% of the length of pWCP and contained no other genomic region. In other words, each of the long reads were equal to or shorter than the length of pWCP, as expected for an extrachromosomal element (Fig. 2a). Moreover, 6 of these 13 reads were exactly equal to the length of pWCP, covering its entirety with non-identical start positions, confirming its circularity (Fig. 2a). The final 6 long reads that covered <50% of the length of pWCP had specific matches to the IS110 TE (Fig. 2b); this result is consistent with the multiple occurrences of the IS110 TE in the Wolbachia genome and explains the increase in coverage in metagenomic read recruitment results (Supplementary Figure 1).

To further validate the extrachromosomal nature of pWCP, we PCR screened genomic DNA from the wild caught individuals as well as from tetracycline (TC)-treated laboratory lines (TC1–TC3; negative controls). TC treatment eliminates Wolbachia from its hosts and is commonly used to generate C. pipiens uninfected laboratory lines\(^{47}\). First, using LCO1490 and HCO2198 mitochondrial cytochrome c oxidase subunit I (COI) invertebrate primers\(^{48}\), we detected the presence of a ~708-bp band in the four C. pipiens individuals and the three Wolbachia-free Culex quinquefasciatus controls treated with TC (TC1–TC3), confirming the presence of arthropod DNA in all samples (Fig. 1b). Next, we verified the presence of Wolbachia DNA in the first four samples by PCR amplifying a 438-bp fragment of the Wolbachia 16S rRNA gene using W-spec-F and W-spec-R primers\(^{49}\). No band was observed in the TC samples, confirming the absence of Wolbachia (Fig. 1c). Critically, a ~1800-bp fragment amplified with primers 263F and 2127R (Supplementary Table 5) designed from the ends of the artificially circularized contig within DnaB gene (and uniquely matching those sites), confirmed the circular nature of the pWCP and its presence only in Wolbachia-infected C. pipiens samples (Fig. 1d). We also designed primers to Sanger sequence across the circular gap (see Supplementary Table 5 for primer sets EC_1–EC_7), and results confirmed the pWCP sequence obtained with both Illumina and MinION sequencing. Finally, we amplified IS110 TE with primers EC_4F and EC_4R (Supplementary Table 5) in the four C. pipiens samples studied herein while observing no band in the Wolbachia-free samples (Fig. 1e) in order to verify the strict association of IS110 TE with Wolbachia (Fig. 1f). These results were further confirmed using additional rifampicin- and oxytetracycline-treated C. pipiens samples (Supplementary Figure 2). Of note, we verified the presence of pWCP in the originating TC1–TC3 and C. pipiens laboratory stocks prior to antibiotic treatment.

The average nucleotide identity of the four independently assembled pWCP sequences from individual mosquitoes was 99.65–100% (Supplementary Table 6). In addition, we identified a 8315-bp contig in the wPip JHB assembly (ABZA01000008)\(^{30}\), which also was >99.53% identical to each of the four pWCP sequences (Supplementary Table 6). The IS110 TE was 100% identical across samples and confirmed as clonal. Our additional read recruitment analyses from publicly available metagenomes (SRR5810516, SRR5810517, SRR5810518)\(^{32}\) also revealed the widespread occurrence of pWCP in C. pipiens individuals from Turkey, Algeria, and Tunisia (Supplementary Note 1, Supplementary Figure 3) at a similar 4.2–7.3-fold higher copy number relative to the Wolbachia genome. We also confirmed that pWCP and the Wolbachia genome display a similar tetranucleotide composition (Supplementary Note 2, Supplementary Figure 4). Overall, these findings suggest that pWCP is maintained over the long term with Wolbachia from these strains.

**pWCP encodes an IS110 TE, 14 genes, and an intergenic repeat region.** The IS element is homologous to ISWpi12 of the IS110 family\(^{46}\) based on IS finder platform search. Annotation of genes in pWCP also revealed the presence of a disrupted DnaB-like helicase, two RelBE loci, a ParA-like gene (each with identical AA sequences across samples, Supplementary Note 3), and multiple genes encoding hypothetical proteins (Fig. 1f, Supplementary Data 3). The ParA-like partitioning sequence showed amino acid homology to the chromosome partitioning of plasmid protein (ParA) identified in Ca. Caedibacter acanthamoebae (an endosymbiont of acanthamoebae; e value: 2 × 10−46), the bacterium Odyssella thessalonicensis (e value: 9 × 10−43), and Rickettsia
raoultii $(e \text{ value: } 7 \times 10^{-41})$. The RelBE toxin–antitoxin (TA) locus has been identified in multiple Wolbachia genomes and is often associated with prophage WO regions (e.g., of $wVitA$, $wHa$, $wMel$, $wAu$, $wRi$, $wSuzi$, $wFol$, $wInc$), specifically within the tail and/or capsid modules. In other bacteria, this TA system can promote the stability of its encoding mobile element, including plasmids or pathogenicity islands, through post-segregational killing of cells that have lost the antitoxin component of the TA operon and a Terminase, respectively (e value $> 0$ in SCOPe, Pfam, and Clusters of Orthologous Group (COG), highlighting the need for further functional characterization of this newly discovered mobile genetic element. In particular, terminase proteins bind and package DNA into the capsids of phage particles. These data indicate that the extrachromosomal DNA could potentially fall into three categories: a simple plasmid, a mini-chromosome of Wolbachia, or a plasmid-like replicon that hijacks the capsids of phage WO.

Beyond the putative coding regions, alignment of all contigs revealed an intergenic variable number tandem repeat (VNTR) region (Fig. 3a, b, Supplementary Figure 5), characterized as 16-nt
Fig. 2 Alignment of MiniON long reads to pWCP. a The alignment of long reads that cover >50% of the pWCP genome (only 12 of 13 total long reads are shown; we omitted 1 from this display solely due to space considerations). Each rectangular figure shows high scoring pairs (HSPs) and their alignments between pWCP and long reads. The broken HSPs that are parallel to each other are due to low-quality regions in long read sequences, and they are shown in different shades. Concentric circles around pWCP demonstrate the alignment of each long read and their start and stop positions. b The alignment of long reads that cover <50% of the pWCP genome (only 5 of 6 are shown). Every long read shown in the figure has a hit to the IS110 TE repeats adjacent to parA that vary in number among individual pWCP sequences. Recent studies observed the same repeat region, identified as pp-hC1A_5, and used it to genotype different strains of Wolbachia\textsuperscript{56,57}, yet these have not been studied at an individual level. The authors suggested that a deletion in the intergenic polymorphic region could serve as a recombination or horizontal gene transfer site\textsuperscript{56}. Alternatively, we hypothesize that the direct repeats, as present in iteron plasmids, could indicate a potential origin of replication and play a role in copy number control\textsuperscript{58}. Our analysis of the pWCP sequence also revealed a
The Wolbachia metapangenome reveals novel viral genes. The assembly and binning of individual mosquitoes from the wild also enabled comparison of the diversity and the gene content of Wolbachia MAGs and wPip Pel in conjunction with the four metagenomes from individual mosquitoes. By linking genes to their abundances in the four metagenomes from individual mosquitoes, we aimed to investigate the diversity and abundance of phage genes in the context of the five previously identified prophage WO regions in wPip Pel (Fig. 4). Some gene clusters within prophage regions appeared to be unique to wPip Pel and were not detected in our metagenomes. It is possible that these genes were not recovered from our MAGs due to small contig size (that is, contigs were too short to be considered for binning). However, our MAGs often carried upstream and downstream phage genes in these regions, suggesting that while some phage genes were conserved across all genomes, others differed significantly from their wPip Pel counterparts (Fig. 4, Supplementary Data 4). It is possible that a set of new phage-associated genes only found in our MAGs are often virus associated^50,61–63. Note that most gene clusters with genes that were unique to MAG O07 did recruit reads from the three other metagenomes (Fig. 4, Supplementary Figure 6), suggesting that, even though they were not characterized in our other MAGs, they did occur in C. pipiens metagenomes. Absence of these genes from our MAGs is most likely due to (1) assembly artifacts that result in fragmented contigs that are too short to be considered for binning or (2) mutations in the gene context that affect the gene caller to identify them properly. Gene clusters that matched to pWCP also occurred only in our Wolbachia MAGs and were missing in wPip Pel (Supplementary Figure 6), which is expected since the wPip Pel reference genome is solely composed of the Wolbachia bacterial chromosome^40. Overall, the metapangenome sheds light on a substantial amount of viral genetic diversity, revealing almost as many virus-associated genes as the ones that were previously recognized in the reference genome wPip.

Unlike our Wolbachia MAGs, wPip Pel is a high-quality genome assembled into a single scaffold. Even though wPip Pel is not completely closed^40, it offers a more complete representation of the synteny of genes in comparison to our MAGs. Hence, in addition to ordering gene clusters based on their distribution patterns across all genomes (Supplementary Figure 6), we took advantage of the wPip Pel genome to determine the order of gene clusters in the metapangenome based on the order of wPip genes in the wPip Pel genome (Fig. 4). This 'forced synteny' allowed us to investigate the diversity and abundance of phage genes in the context of the five previously identified prophage WO regions in wPip Pel (Fig. 4). Some gene clusters within prophage regions appeared to be unique to wPip Pel and were not detected in our metagenomes. It is possible that these genes were not recovered from our MAGs due to small contig size (that is, contigs were too short to be considered for binning). However, our MAGs often carried upstream and downstream phage genes in these regions, suggesting that while some phage genes were conserved across all genomes, others differed significantly from their wPip Pel counterparts (Fig. 4, Supplementary Data 4). It is possible that a set of new phage-associated genes only found in MAGs (Fig. 4, Supplementary Data 5) have functional homologues in wPip Pel. Previous studies indeed show WO genes that have distinct nucleotide sequences yet similar predicted functions^30. However, it is also possible that some genes detected only in our MAGs at the sequence level may also be encoding unique functions compared to known phage genes; for instance, eukaryotic-like homologues were recently shown as constituents of phage WO^27.
Metagenomic read recruitment revealed between a 1.5- and 5-fold increase in coverage between pWCP and some structural phage genes (e.g., WP0415 and WP0446 Tail genes) in our metagenomes compared to the coverage of the bacterial chromosome in all four *C. pipiens* individuals (Supplementary Data 6). The forced synteny organization of gene clusters also revealed that a single prophage WO region could include both high- and low-coverage phage genes (Fig. 4). The multi-copy occurrence of pWCP could explain its increased coverage (Supplementary Figure 1), and differential coverage regimes for genes within a single prophage region could be explained by at least two different models. First, increased coverage of some prophage genes could be attributed to lytic activity: the prophage genes displaying lower bacterial-like coverage are not part of the virion, while those with higher coverage correspond to phage genes that are replicated and packaged into phage capsids. This lytic model is consistent with the observation of phage particles in *C. pipiens* mosquitoes44,45, observed lytic activity of *Wolbachia* phages in *Nasonia* testes, and the sequencing of WO genomes from purified phage particles27. The partial replication and packaging of prophage WO genes could result from either a ‘less than headful’ mechanism of packaging, as described in model
phages P1 and \textsuperscript{T}4\textsuperscript{6,65}, or it could represent active vs. degenerate prophage variants in the \textit{w}Pip \textit{P}el chromosome. Second, some genes in prophage \textit{WO} genomes could be copied throughout the \textit{Wolbachia} chromosome explaining the increase in coverage due to their multi-copy occurrence. This model is supported by the prophage duplication events observed in the \textit{w}Ri and \textit{w}Suz genomes\textsuperscript{61,66} as well as the presence of TEs within and/or flanking prophage variants\textsuperscript{27,40} that could enable genomic rearrangement and duplication. These models may not be mutually exclusive, and the system could involve both duplication of prophage genomes and the induction of phage particles.

**Discussion**

Shotgun metagenomes from individual \textit{C. p. ovary} samples allowed us to de novo reconstruct \textit{Wolbachia} genomes from single mosquitoes and compare these MAGs to each other as well as to the reference \textit{w}Pip \textit{P}el genome through pangenic strategies. Our data reveal an extensive diversity of previously undetected \textit{Wolbachia} phage \textit{WO} and other viral genes and notably the first indication that \textit{Wolbachia} harbours a candidate plasmid, shedding new light on the richness of the \textit{Wolbachia} mobilome. The definition of a plasmid varies; here we adhere to the typical characterization of a plasmid as a small hereditary, extrachromosomal, circular element.

Even though a \textit{Wolbachia} plasmid has not been reported before, we did find evidence of its occurrence in \textit{C. p.} metagenomes in previous studies. It is likely that the previous efforts overlooked this element due to computational challenges associated with the assembly of metagenomic short reads. Beyond the IS110 TE that occurred both in the plasmid and \textit{Wolbachia} chromosome, \textit{pWCP} contained a region of intergenic VNTR that differed across individuals in our study. This suggests that the co-assembly of pooled individuals may yield fragmented contigs. The same VNTR sequences were found in \textit{C. p.} from Australia, Argentina, USA, Italy, Japan, Israel, and Greece\textsuperscript{56}, and we confirmed the presence of \textit{pWCP} in \textit{C. p.} samples from countries of the Mediterranean basin\textsuperscript{32}. The high similarity of \textit{pWCP} sequences across \textit{C. p.} metagenomes in addition to its global prevalence suggest evolutionary constraints and a possible functional role in \textit{Wolbachia} symbiosis. However, we did not detect the plasmid in other \textit{Wolbachia} strains through screening available metagenomes from the \textit{Ily Drosophila melano- gaster}, the planthopper \textit{Laodelphax striatella}, and \textit{Anopheles gambiæ} mosquito.

Our work demonstrates that the combination of genome-resolved metagenomics, long-read sequencing, and pangenic strategies provide an effective computational framework to investigate the diversity and distribution of mobile genetic elements in endosymbionts that are challenging to cultivate. Furthermore, it shows the importance of studying distinct individuals from wild mosquito populations, in parallel with controlled experiments in laboratory settings, to improve our understanding of the \textit{Wolbachia} mobilome. The fragmented nature of \textit{Wolbachia} MAGs in our study emphasizes the critical need for harnessing the power of emerging long-read sequencing technologies to characterize complex genomic variations of \textit{Wolbachia} and its mobilome at finer scales. \textit{Wolbachia} has been so far recalcitrant to genetic modification, but the discovery of \textit{pWCP} and \textit{phage \textit{WO}} may create new avenues for effective genome-editing strategies.

**Methods**

**Sample collection.** We collected mosquito specimens using a carbon dioxide mosquito trap located in Languedoc, Hérault, France in May 2017 (Camping l’Europe de Vic La Gardiole, EID Méditerranée). Specimens were transported alive to the laboratory immediately upon recovery. We anesthetized adult females for 4 min at −20 °C and proceeded to species-level identification. To remove potential contaminants from the insect surface, we gently vortexed specimens for 1 min in Luedli (4 °C) 96% ethanol. We transferred their ovaries to 1.2 ml vials with 1 ml sterile cold (4 °C) phosphate-buffered saline (PBS) 1x solution and gently vortexed them again for 10 s to avoid DNA precipitation with ethanol. Finally, we transferred specimens onto a sterile microscope slide with sterile PBS 1x on top of a cold plate and dissected two ovaries from four specimens using sterilized tweezers. We preserved ovary samples at −80 °C until further processing.

**C. p. complex controls.** We obtained \textit{Wolbachia} uninfected mosquitoes by treating the host with antibiotics. For antibiotics treatment we either used TC, which inhibits protein synthesis, or rifampicin, which interferes with nucleic acid synthesis\textsuperscript{67}. The TC-treated lines\textsuperscript{67} shown in Fig. 1 \textit{(C. quinquefasciatus SLAB-TC lines, ISEM, France)} were raised at least 1 year without \textit{TC} in standard laboratory conditions before beginning experiments. The rifampicin and oxytetracycline-treated \textit{C. p. lines} were kindly provided by Maria del Mar Fernandez de Marco (Animals Plant Health Agency, UK).

**Metagenomic library preparation and sequencing.** We extracted total genomic DNA from each ovary sample, hereafter referred to as O03, O07, O11, and O12, using the MoBio PowerFecal DNA Isolation Kit (QIAGEN Inc., Germantown, MD, USA). We used an E220 Covaris instrument (Covaris, Woburn, MA, USA) to sonicate 3.8-5.7 ng of genomic DNA. We end-removed and 3′-adenylated resulting fragments and used the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) to add NEXTflex PCR-free barcode adapters (Bioro Scientific, Austin, TX, USA). We purified ligation products by Beckman Coulter, Brea, CA, USA) and PCR-amplified fragment libraries (200 bp; 2 PCR reactions, 14 cycles) using Illumina adapter-specific primers and NEBNext Ultra II Q5 Master Mix (NEB). After library profile analysis using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and quantitative PCR quantification using the KAPA Library Quantification Kit for Illumina Libraries (KapaBiosystems, Wilmington, MA, USA), we sequenced the library using a HiSeq4000 Illumina sequencer (Illumina, San Diego, CA, USA) at the Genoscope in Evry, France, generating 151 bp paired-end reads. To remove the least reliable data, we filtered the raw sequencing results using cluster intensity and chiasity filter as described in ref.\textsuperscript{68}. We used illumina-utils v1.4.4\textsuperscript{69} to quality-filter raw paired-end reads with the program ‘in-filter-quality-mimeo’ with default parameters, IDBA-UD v1.1.2\textsuperscript{70} to assemble paired-end reads, and Bowtie2 v2.2.9\textsuperscript{71} for all read alignment analyses. We processed our contigs that are >1000 nts and read recruitment results with anvi\textsuperscript{v}5\textsuperscript{72} to recover MAGs from \textit{C. p.} metagenomes. Briefly, we used the program ‘anvi-gen-contigs-database’ to generate anvi\textsuperscript{v}0 contigs databases for each four individual assemblies, during which anvi\textsuperscript{v}0 calculates and stores tetratrdode frequency values and Prodigal v2.6.3\textsuperscript{33} identifies ORFs in each contig. We ran default anvi\textsuperscript{v}0 HMM profiles on resulting contigs databases using the program ‘anvi-run-hmms’ and assigned functions to genes using the program ‘anvi-run-ncbi-cogs’, which searches gene amino acid sequences against the December 2014 release of the COG data-base. We first blastp v2.3.0-\textsuperscript{v5}\textsuperscript{73} for each contig recruited from each four individual metagenomes onto the four assemblies using the program ‘anvi-profile’, which generates anvi\textsuperscript{v}0 profile databases that store the coverage and detection statistics of each contig within each sample independently. We then merged resulting anvi\textsuperscript{v}0 profile databases for each sample into the program ‘anvi-merge’. For an initial coarse binning, we used the CONCOCT\textsuperscript{v} algorithm through the program ‘anvi-cluster-with-concoct’ and confined the number of clusters to five using the parameter ‘–num-clusters-requested 5’. We then used the program ‘anvi-script-get-collection-info’ to identify bins with a bacterial population genome, manually refined them to identify bacterial genomes using the program ‘anvi-refine’, and assigned taxonomy to resulting MAGs using NCBI’s non-redundant protein sequence database with amino acid sequences of core genes. The program ‘anvi-refine’ allows the identification and refinement of population genome bins through an interactive interface by offering a number of tools including (1) guiding hierarchical clustering dendrograms of contigs (including one based on tetratetode frequencies and differential coverage statistics across samples), (2) real-time completion and redundancy estimates of binned contigs based on bacterial single-copy core genes\textsuperscript{37}, (3) interfaces that display nucleotide-level coverage statistics per gene and contig, (4) functional annotation and synteny of genomes, and (5) online search options for gene and contig sequences. These tools collectively help minimize the inclusion of non-target contigs (such as eukaryotic contamination) in final bins, especially in complex metagenomes\textsuperscript{37}. Except for the manual refinement step, we automated the processing of our raw sequencing data with a snakemake\textsuperscript{workflow} anvi\textsuperscript{v}0 v5\textsuperscript{55} (http://meerenlab.org/2018/07/09/anvi-0-snakeake-workflows/), and the Code availability section reports necessary configuration files to fully reproduce this analysis.

**Sample preparation and long-read sequencing.** To investigate the circularity of \textit{pWCP}, we prepared for a long-read sequencing strategy. We extracted and pooled

---

**ARTICLE**

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-08973-w | www.nature.com/naturecommunications
DNA from the ovaries of 7 C. pipiens and 15 C. quinquefasciatus mosquitoes. As the total DNA quantity from this step was only 5.83 ng, we also extracted DNA from the abdomens of 15 C. quinquefasciatus that left us with a total of 76.63 ng. To acquire high-molecular weight DNA from these low-biomass samples, we used an in-house Phenol Chloroform protocol for total genomic DNA extraction. Briefly, we used a chloridic acid-washed pestle adapted to the size of a microcentrifuge tube to gently grind tissue in cetyltrimethylammonium bromide buffer (10 mL) with betamercaptoethanol (20 μL) that we incubated for 15 min at 60 °C. We then added 200 μL of phenol/chloroform/isoamyl alcohol and incubated each microcentrifuge tube at room temperature while rotation shaking for 5 min. To minimize inclusion of proteins, we centrifuged the tubes for 10 min at 8000 rpm (9393 × g), removed the aqueous phase (the top layer), and placed it into fresh tubes. We repeated those steps with chloroform:isoamyl alcohol and lastly with isopropanol. We incubated microcentrifuge tubes for 10 min at room temperature and mixed by inversion every 3 min. We then centrifuged those tubes for 10 min at 10,000 rpm (9279 × g), removed the top layer, and kept the sole DNA at the bottom of the microcentrifuge tube. We next washed DNA with 400 μL of freshly prepared 70% ethanol and centrifuged tubes for 10 min at 10,000 rpm with 20 μL of molecular grade water. During all laboratory steps, we performed smooth and slow pipetting with cut pipette tips to avoid further shearing the DNA molecules. To prepare mosquito total DNA for sequencing, we used the Rapid Barcoding Kit (SQK-RBK004) by Oxford Nanopore Technologies. We followed the manufacturer protocols, except for one additional step, in which we supplemented samples that did not meet the manufacturer DNA input mass recommendations (400 ng) with linear double-stranded lambda DNA (New England Biolabs) for “padding.” We then used 1× Agencourt AMPure XP beads (A63882, Beckman Coulter) to sample bead and concentration of pooled barcoded samples. A MinION with a pristine R9/4/FLO-MIN106 flow cell (Oxford Nanopore Technologies) sequenced the final prepared library with a starting voltage of ~180 mV and a run time of approximately 20 h. While the flow cell had 1293 active pores prior to the run, only 200 were active at the beginning of the run, dropping down to 2 active pores by the end of the run. We were not able to determine whether the low number of initial pores and their rapid decline was due to remaining chemical residues from our in-house Phenol Chloroform extraction protocol or a defective flow cell. We used MinKNOW (v1.15.4) for live base-calling and Albacore (v2.1.7) to de-multiplex our samples and convert raw FAST5 files to FASTQ files. For downstream analyses, we only used reads with a minimum quality score of 7.

Pangenomic analysis and the metapangenome. Our analysis of the Wolbachia pangenome followed the workflow outlined in ref. 86 and at http://merenlab.org/2016/11/08/pangenomics-v2/. Briefly, we first generated an avi’o genomes storage database using the program ‘avi’o-genomes-storage’ for our four Wolbachia MAGS and the wpPip reference genome wpPip PeL (NCBI Accession ID NC_010981.1). We then used the program ‘avi’o-pan-genome’ using the flag ‘--use-uchi-blast’, and parameters ‘--minhit 0.5’, and ‘--mcl inflation 5’. This program (1) uses the blastp program all vs. all to create a graph of similarities between pairs of amino acid sequences, (2) removes bad hits using the ncbi-blast platform without any filters (following the manufacturer protocols, except for one additional step, in which we supplemented samples that did not meet the manufacturer DNA input mass recommendations (400 ng) with linear double-stranded lambda DNA (New England Biolabs) for ‘padding’). We then used 1× Agencourt AMPure XP beads (A63882, Beckman Coulter) to sample bead and concentration of pooled barcoded samples. A MinION with a pristine R9/4/FLO-MIN106 flow cell (Oxford Nanopore Technologies) sequenced the final prepared library with a starting voltage of ~180 mV and a run time of approximately 20 h. While the flow cell had 1293 active pores prior to the run, only 200 were active at the beginning of the run, dropping down to 2 active pores by the end of the run. We were not able to determine whether the low number of initial pores and their rapid decline was due to remaining chemical residues from our in-house Phenol Chloroform extraction protocol or a defective flow cell. We used MinKNOW (v1.15.4) for live base-calling and Albacore (v2.1.7) to de-multiplex our samples and convert raw FAST5 files to FASTQ files. For downstream analyses, we only used reads with a minimum quality score of 7.

PCR amplification of individuals. We used a set of specific primers on the four C. pipiens individuals and three Wolbachia-free C. quinquefasciatus controls generated from total genomic DNA using a 438-bp fragment of the Wolbachia 16S rDNA gene using Wspec-F: TAA ACT TGA ATG TAT GGC and Wspec-R: TTA AAC TTA GGG TCA CAA AAA AAT CA -3’. 88 We checked for the presence of Wolbachia DNA by PCR targeting the following genes: 16S rDNA (5’–3’ TAA ACT TGA ATG TAT GGC TTA AAC TTA GGG TCA CAA AAA AAT CA -3’).” 53

Plasmid characterization. We used the Glimmer software within Geneious v8.1.9 (Biomatters, Ltd) to identify ORFs and used the NCBI databases for non-redundant protein sequences and conserved domains54, SMART55, and HHpred84 (including SCOPe, Pfam, and COG) to manually annotate putative functions based on amino acid sequence homology searches. We used the Is finder platform (https://isfinder.biotoul.fr/) for IS homology search. The nucleotide analysis plugin in Geneious for EMBOSS identified and illustrated the EP in the plasmid.

Figure. We visualized BLAST hits between pWCP and long read sequences using Kablammo85 (http://kablammo.wasmuthlab.org/) and finalized all figures using the open-source vector graphics editor Inkscape (https://inkscape.org/).

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The raw sequencing data for shotgun metagenomes are available in the European Nucleotide Archive via accession code PRJEB26028. Sequencing data for the MinION run is available at https://doi.org/10.6084/m9.figshare.7306784. We also made available FASTA files for individual metagenomic assemblies (https://doi.org/10.6084/m9.figshare.6263867), the four metagenome-assembled Wolbachia genomes (https://doi.org/10.6084/m9.figshare.6293040), artificially circularized individual plasmid sequences (https://doi.org/10.6084/m9.figshare.6380015), as well as avi‘o merged profile databases (https://doi.org/10.6084/m9.figshare.6263876) and avi‘o files for the Wolbachia pangenome (https://doi.org/10.6084/m9.figshare.6291650).

Received: 10 August 2018 Accepted: 6 February 2019 Published online: 05 March 2019

References
1. Gould, E., Pettersson, J., Higgs, S., Charrel, R. & de Lamballerie, X. Emerging mosquito vectors of Zika virus. Nat. Rev. Microbiol. 16, 508–518 (2018).
2. Boyer, S., Calvez, E., Chouin-Carneiro, T., Diao, D. & Failloux, A. B. An overview of mosquito vectors of Zika virus. Microbes Infect. 20, 646–660 (2018).
3. Flores, H. A. & O'Neill, S. L. Controlling vector-borne diseases by releasing modified mosquitoes. Nat. Rev. Microbiol. 16, 508–518 (2018).
4. Taylor, M. J., Bordenstein, S. R. & Statko, B. Microbe profile: Wolbachia: a sex detector, a viral protector and a target to treat filarial nematodes. Microbiology 164, 1345–1347 (2018).
5. LePage, D. & Bordenstein, S. R. Wolbachia: can we save lives with a great pandemic? Trends Parasitol. 29, 385–393 (2013).
6. Stouthamer, R., Breeuwer, J. A. J. & Hurst, G. D. Wolbachia pipientis: microbial manipulator of arthropod reproduction. Annu. Rev. Microbiol. 53, 71–102 (1999).
27. Bordenstein, S. R. & Bordenstein, S. R. Eukaryotic association module in Wolbachia induces resistance to RNA viral infections in Drosophila melanogaster. *PLoS Biol.* 6, 2753–2763 (2008).

26. LePage, D. P. et al. Prophage WO genes recapitulate and enhance cytoplasmic incompatibility in *Aedes aegypti*. *PLoS Negl. Trop. Dis.* 11, 1–14 (2017).

25. Beckmann, J. F., Ronau, J. A. & Hochstrasser, M. A. A deubiquitylating enzyme induces cytoplasmic incompatibility and blocks dengue transmission in *Aedes albopticus*. *Proc. Natl Acad. Sci. U.S.A.* 109, 255–260 (2012).

24. McMeniman, C. J. et al. Stable introduction of a life-shortening Wolbachia infection into the mosquito *Aedes aegypti*. *Sci. Rep.* 3, 1441–1449 (2014).

23. Atyame, C. M. et al. Cytoplasmic incompatibility as a means of controlling mosquito vectors. *Nature* 436, 20–24 (2005).

22. Funkhouser-Jones, L. J., van Opstal, E. J., Sharma, A. & Bordenstein, S. R. The ultrastructure of the rickettsia-like microorganisms in Wolbachia patients and associated virus-like bodies in the mosquito *Culex quinquesetus*. *J. Ultrastruct. Mol. Biol.* 70, 13–18 (1964).

21. Werren, J. H. Biology of *Wolbachia*. *Annu. Rev. Entomol.* 124, 587–609 (1997).

20. Atyame, C. M. et al. Wolbachia divergence and the evolution of cytoplasmic incompatibility in *Culex pipiens*. *PLoS One* 9, 21–26 (2014).

19. Atyame, C. M. et al. Cytoplasmic incompatibility as a means of controlling *Culex pipiens quinquesetus* mosquito in the islands of the South-Western Indian Ocean. *PLoS Negl. Trop. Dis.* 5, 20–22 (2011).

18. Serbus, L. R. & Sullivan, W. A. Cellular basis for *Wolbachia* recruitment to the host germline. *PLoS Pathog.* 3, 1930–1937 (2007).

17. Funkhouser-Jones, L. J., van Opstal, E. J., Sharma, A. & Bordenstein, S. R. The maternal effect gene Wd controls *Wolbachia* titer in *Nasonia*. *Curr. Biol.* 28, 1692–1702.e6 (2018).

16. Kent, B. N. et al. Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*) in *Drosophila melanogaster*. *PLoS Pathog.* 1, 301–312 (2005).

15. McMeniman, C. J. et al. Harnessing mosquito-*Wolbachia* symbiosis for vector and disease control. *Acta Trop.* 132, 5150–5163 (2014).

14. Mains, J. W., Brelsfoard, C. L., Rose, R. I. & Dobson, S. L. Female adult localization in the host germline. *Mol. Ecol.* 1175–1178 (2011).

13. Baldridge, G. D., Burkhardt, N. Y., Felsheim, R. F., Kurtti, T. J. & Luderer, U. G. Transposon insertion reveals PRM, a plasmid of *Rickettsia monacensis*. *Appl. Environ. Microbiol.* 73, 4984–4995 (2007).

12. Ogata, H. et al. The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biol.* 3, e292 (2005).

11. Bonneau, M. et al. *Culex pipiens* crossing type diversity is governed by an amplified and polymorphic operon of *Wolbachia*. *Nat. Commun.* 9, 319 (2018).

10. El Karkouri, K., Pontarotti, P., Raoult, D. & Fournier, P. E. Origin and evolution of rickettsial plasmids. *PLoS ONE* 11, 1–17 (2016).

9. Baldridge, G. D., Burkhardt, N. Y., Felsheim, R. F., Kurtti, T. J. & Luderer, U. G. Transposon insertion reveals PRM, a plasmid of *Rickettsia monacensis*. *Appl. Environ. Microbiol.* 73, 4984–4995 (2007).

8. Hedges, L. M., Brownlie, J. C., O’Neill, S. L. & Johnson, K. N. Wolbachia and virus protection in insects. *Science* 322, 1489–1491 (2008).

7. Tenuta, L., Fernández, M. F. & Aebisher, M. The bacterial symbiont *Wolbachia* induces resistance to DNA viral infections in *Drosophila melanogaster*. *PLoS Biol.* 6, 2753–2763 (2008).

6. Jeffries, C. L. & Walker, T. *Wolbachia* biocost control strategies for arboviral diseases and the potential influence of resident *Wolbachia* strains in *Aedes albopictus*. *Mosq. News*. 75, 20–25 (2015).

5. Joubert, D. A. & O’Neill, S. L. Comparison of stable and transient *Wolbachia* infection models in *Aedes aegypti* to block Dengue and West Nile viruses. *PLoS Negl. Trop. Dis.* 11, 1–14 (2017).

4. Klasson, L. et al. Genome evolution of *Wolbachia* strain wPip from the mosquito *Aedes albopictus* strain MelPop. *PLoS Pathog.* 10, e1004122 (2014).

3. Iurbe-Ormaetxe, I., Woolf, M., Rancés, E., Duplouy, A. & O’Neill, S. L. A simple protocol to obtain highly pure *Wolbachia* endosymbiont DNA for genome sequencing. *J. Microbiol. Methods* 84, 134–136 (2011).

2. Wyke, J. D., Sjostrand, S. F., Portaro, J. K. & Barr, A. R. The ultrastructure of the rickettsia-like microorganism in *Wolbachia* patients and associated virus-like bodies in the mosquito *Culex quinquesetus*. *J. Ultrastruct. Mol. Biol.* 70, 13–18 (1964).

1. Cervenka, N., Leclercq, S., Leroy, E., Bouchon, D. & Cordaux, R. Short- and long-term evolutionary dynamics of bacterial insertion sequences: Insights from *Wolbachia* endosymbionts. *Genome Biol. Evol.* 3, 1175–1186 (2011).
Acknowledgements

We thank Albane Marie, Gregory Lambert, and Jean-Baptiste Panchau from EID (Entente Interdépartementale de Démosuscitation) for field Culex pipiens samples; Maria del Mar Fernandez de Marco from Animals Plant Health Agency for antibiotic-treated Culex pipiens individuals; Evan Kieffer, Richard Fox, and Frederick Gavory for their help with computational matters; Phoebee Rice, Sean Crosson, and Marie-Agnès Petit for their insights into sequencing. This research was funded by the National Institute for Agronomical Research (INRA) Young Researcher grant (2017) to J.R. and National Institutes of Health awards R01 AI132581 and R21AI133522, National Science Foundation award 145677, and Vanderbilt Microbiome Initiative funding to Seth R Bordenstein. The INRA Metaprogram (Microbial Ecology and Metagenomics) MEM from INRA provided a USA mobility grant to J.R., and A.M.E. was funded by the University of Chicago start-up funds and the Keck Foundation. K.L. was supported by a Gastrointestinal Research Foundation grant. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

J.R. conceived the study, performed laboratory assays, analysed the data, prepared figures and tables, and wrote the manuscript; Sarah R. Bordenstein analysed the data, initially identified pWCP, performed laboratory assays, prepared figures and tables, and wrote the manuscript; C.C., M.W., P.M., and J.R. performed laboratory assays; A.S. and O.C.E. provided data analysis tools; K.L. and A.R.W. performed biological assays and long-read sequencing; Seth R. Bordenstein supervised the research and wrote the manuscript; A.M.E. provided data analysis tools, supervised the research, analysed the data, prepared figures and tables, and wrote the manuscript. All authors edited and approved the final draft.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-08973-w.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: Nature Communications thanks Andreas Brenges, Lisa Klasson and the other anonymous reviewer for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019