PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Martinez J, Bruner-Montero G, Arunkumar R, Smith SCL, Day JP, Longdon B, Jiggins FM. 2019 Virus evolution in *Wolbachia*-infected *Drosophila*. *Proc. R. Soc. B* **286**: 20192117.
http://dx.doi.org/10.1098/rspb.2019.2117

Received: 12 September 2019
Accepted: 8 October 2019

Subject Category:
Evolution

Subject Areas:
evolution, genetics

Keywords:
Wolbachia, *Drosophila melanogaster*, antiviral resistance, experimental evolution

Author for correspondence:
Julien Martinez
e-mail: julien.martinez@glasgow.ac.uk

†Present address: MRC-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow G61 1QH, UK.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.4698191.

THE ROYAL SOCIETY PUBLISHING

Virus evolution in *Wolbachia*-infected *Drosophila*

Julien Martinez[1,†], Gaspar Bruner-Montero[1], Ramesh Arunkumar[1], Sophia C. L. Smith[1], Jonathan P. Day[1], Ben Longdon[1,2] and Francis M. Jiggins[1]

[1]Department of Genetics, University of Cambridge, Cambridge, UK
[2]Centre for Ecology and Conservation, University of Exeter, Penryn Campus, Cornwall TR10 9FE, UK

JM, 0000-0001-8210-2921; GB-M, 0000-0002-4614-0338; BL, 0000-0001-6936-1697; FMJ, 0000-0001-7470-8157

Wolbachia, a common vertically transmitted symbiont, can protect insects against viral infection and prevent mosquitoes from transmitting viral pathogens. For this reason, *Wolbachia*-infected mosquitoes are being released to prevent the transmission of dengue and other arboviruses. An important question for the long-term success of these programmes is whether viruses can evolve to escape the antiviral effects of *Wolbachia*. We have found that *Wolbachia* altered the outcome of competition between strains of the DCV virus in *Drosophila*. However, *Wolbachia* still effectively blocked the virus genotypes that were favoured in the presence of the symbiont. We conclude that *Wolbachia* did cause an evolutionary response in viruses, but this has little or no impact on the effectiveness of virus blocking.

1. Introduction

Wolbachia is a maternally transmitted intracellular bacterium found in many insects [1]. Its ability to spread rapidly through insect populations by inducing a sperm–egg incompatibility called cytoplasmic incompatibility [2–4], coupled with its inhibitory effect on the replication of RNA viruses [5–8], make it a promising control agent to prevent the transmission of mosquito-borne diseases [9]. In several parts of the world, the bacterial symbiont is being introduced into natural populations of the mosquito *Aedes aegypti*, the main vector of dengue and Zika viruses [10–13]. Preliminary field releases of *Wolbachia*-infected *Ae. aegypti* females have demonstrated that the bacterial infection is able to spread and be stably maintained [4,10,11], turning susceptible populations of mosquitoes into virus-resistant ones [14].

Like other control methods, there is a risk that the release of *Wolbachia*-infected mosquitoes may not be evolution-proof. It may promote adaptive changes in the mosquito vector, *Wolbachia* or virus that could hamper the long-term success of field interventions. Therefore, there is an urgent need to understand and predict what genetic changes might follow the introduction of *Wolbachia*, especially because such introductions are likely to be irreversible [15,16]. For instance, high antiviral resistance is associated with high densities of the symbiont within the insect tissues [17,18], and this leads to reductions in the fecundity, lifespan and other fitness-related traits of the insect host [11,19–22]. These costs may lead to the evolution of lower *Wolbachia* densities and thus a reduction or loss of the antiviral phenotype. A second concern is the evolution of the virus itself. Since *Wolbachia* blocks the transmission of the virus by inhibiting its replication, virus populations should be selected to overcome such inhibition. Potentially, virus strains that are able to replicate at a higher rate in the presence of *Wolbachia* could be advantaged and spread. For example, the intensity of *Wolbachia*'s effect on dengue transmission varies between virus serotypes [14], and the magnitude of these differences is sufficiently large that it is predicted to alter the outcome of control programmes [23]. Therefore, if viruses can escape the

resistance conferred by *Wolbachia*, this would threaten the sustainability of symbiont-based interventions.

Here, we passaged Drosophila C virus (DCV) through *Wolbachia*-infected *Drosophila melanogaster* and examined how the symbiont affected the evolution of the virus. DCV is a positive-strand RNA virus of the family Discistroviridae that naturally infects *D. melanogaster* [24]. It is highly pathogenic in laboratory experiments, leading to fly death within a few days [25]. *Wolbachia* leads to large reductions in DCV titres and increases survival after DCV infection, which has led to DCV becoming a common model to study *Wolbachia*-mediated antiviral protection [5,17,18]. We found that populations of the virus became genetically differentiated from controls in *Wolbachia*-free flies, with the same viral genotype being favoured across replicate populations. However, despite these parallel genetic changes providing evidence of adaptive evolution in the viral populations, we could not detect any reduction in *Wolbachia*'s antiviral effect or any increase in DCV virulence.

2. Methods

(a) Fly husbandry and virus isolates

Drosophila melanogaster lines previously described in [21] were kindly provided by Luis Teixeira. The DrosDel w^{1118} isogenic background was used as the *Wolbachia*-free control. The *Wolbachia*-infected line was created in [21] by introgressing the DrosDel w^{1118} nuclear background into a cytoplasm infected with the *Wolbachia* strain wMelCS_b through chromosome replacement using balancers for the first, second and third chromosomes (the fourth chromosome was not replaced). Flies were maintained on a cornmeal diet (1200 ml water, 13 g agar, 105 g dextrose, 105 g maize, 23 g yeast, 35 ml Nipagin) at 25°C under a 12 h light–dark cycle and 70% humidity.

DCV isolates were previously described in [26] and kindly provided by Karyn N. Johnson. Isolates DCV-C and -G originate from France, DCV-EB and -CYG from Australia, and DCV-M, -O, -T and -Z from Morocco.

(b) Virus production

All DCV isolates were passaged once in Schneider Drosophila Line cells (DL2) before the series of experiments. Cells were cultured at 26.5°C in Schneider's Drosophila medium with 10% fetal bovine serum, 100 U ml^{-1} penicillin and 100 mg ml^{-1} streptomycin (all Invitrogen, UK). Cells were then freeze–thawed twice to lyse cells and centrifuged at 4000*g* for 10 min at 4°C to remove cellular debris. Finally, the supernatant containing DCV was aliquoted and frozen at −80°C.

For infection assays, aliquots of virus solutions were defrosted on the day of infection. Virus aliquots were diluted in Ringer's solution [27] to standardize the concentration of DCV RNA measured by quantitative PCR (qPCR, see DCV titre for primers and amplification cycles). The concentration of DCV RNA was used instead of the TCID50 method [28] as the cytopathic effects of the eight DCV isolates differed considerably. Total RNA from the eight virus solutions was extracted and a standard *Wolbachia* spike-in added during the RNA extraction. The DCV critical threshold values were then normalized relative to the *Wolbachia* gene *atpD* using primers atpDQALL_F (5′-CCTTATCTTAAAGGAGGAAA-3′) and atpDQALL_R (5′-AATCCTTTATGAGCTTTTGC-3′).

(c) Sequencing of Drosophila C virus genome and phylogenetic analysis

Viral RNA was extracted using TRIzol reagent (Invitrogen) and reverse-transcribed with the Maxima H Minus Reverse Transcriptase (ThermoFisher Scientific) and oligo dT primers. The DCV genome was then amplified by PCR using a set of 22 primers distributed along the 9264 bp genome (electronic supplementary material, table S1). For each genome, the 11 PCR products were Sanger-sequenced and the reads assembled into a consensus genomic sequence using Sequencher v. 4.5 software (GenBank accession numbers: MK645238–MK645245). DCV genomes were aligned with ClustalW in BioEdit v. 7.0.9 [29]. A maximum-likelihood phylogenetic tree was built using the aligned genomes and node supports were assessed with 1000 bootstrap replications.

(d) Selection experiment 1 on genetically diverse virus populations

To investigate virus adaptation from standing genetic variation, we created a virus population composed of equal proportions of each of the eight DCV isolates based on the relative amount of DCV RNA in the original virus solutions. The DCV mixture was then passaged in flies with or without *Wolbachia* by infecting 3–6-day-old female flies (passage 1). For this, flies were anaesthetized on CO_2 and stabbed on one side of the thorax with a steel needle (Austerlitz Insect Pins) dipped into the virus solution as explained in [30]. Ten flies were then transferred to a vial containing cornmeal food. Twelve vials per *Wolbachia* treatment were prepared, representing 12 biological replicates. Virus populations were harvested 3 days post-infection by homogenizing the 10 virus-infected flies from each vial in 25 µl of Ringer's solution. The homogenate was then centrifuged at 12 000*g* and 10 µl of the supernatant was frozen at −80°C and later used as the inoculum for further passage. The remaining 20 µl containing the fly tissues were diluted in 250 µl of TRIzol reagent and frozen at −80°C for later RNA extraction. The virus populations were serially passaged two more times (passages 2 and 3) by repeating the steps above and infecting new flies from the respective *Wolbachia* treatment. Note the amount of virus in the inoculum was only controlled at the start of selection (passage 1). In order to avoid cross-contamination between virus populations, different tools (needles, handling brushes) were used for each replicate population at each passage.

(e) Pool-sequencing, read processing, mapping and variant identification

Total RNA was extracted from flies in passage 3 of the selection experiment 1. Since the DCV genome contains a poly(A) tail, we isolated the virus genomic RNA along with fly transcripts by capturing the polyadenylated RNAs from the extracted total RNA using the KAPA Stranded mRNA-Seq kit (Kapa Biosystems) and a different indexed adapter for each RNA library. Twenty-four libraries were prepared (one per virus population), quantified by qPCR and pooled in equal proportions into a multiplexed library. The pool was sequenced in one lane of Illumina HiSeq4000 to generate single-end 50 bp reads (SRA study accession number PRJEB21984).

We used Trimmomatic v. 0.32 [31] to trim reads. We first removed three bases from the 3′ end of the read. Reads were quality trimmed from the 3′ end, cutting when the average quality scores in sliding windows of four bases dropped below 15. We required reads have a minimum length of 36. Using *BWA MEM* [32], we mapped reads to the genome of a genetically homogeneous DCV population (DCV-ref) previously produced from the isolate DCV-C by endpoint dilution [33]. We removed optical duplicate reads using Picard tools (https://broadinstitute.github.io/picard/). We realigned reads close to indels using GATK [34]. Following that, we used Samtools [35] to remove reads with mapping quality scores lower than 40. We also used Samtools to generate a multi-pileup file to assist with variant

identification. Finally, we used PoPoolation2 to identify the allele frequency of the variants in the pooled genomic library samples [36]. To make the statistical power and influence of different variants and libraries similar, we down-sampled sites to a maximum coverage of 50×/library.

(f) Recombination analysis

In order to investigate the extent of recombination of the DCV genome in selection experiment 1, we resequenced the 24 RNA libraries using MiSeq to generate paired-end 250 bp reads (SRA study accession number PRJEB21984). The software bcl2fastq on BaseSpace (Illumina Inc., San Diego, CA, USA) was used to demultiplex the base call files to Fastq files, trim adaptor sequences using a sliding window with an adapter stringency of 0.9. We then used Trimmomatic [31] to cut reads at the first base where the quality score (Q) dropped below 30, retaining only reads with a minimum length of 35 bp. These were mapped to the genome of isolate DCV-C (GenBank accession number MK645242) using *BWA MEM* [32]. We used Picard tools (https://broadinstitute.github.io/picard/) to reorder reads, add read groups and sort them by coordinates. Mapped reads were converted into a table of variants where each row is a separate sequence read using sam2tsv in JVarkit [37]. The reads from all libraries were combined. We kept properly paired reads, as identified by their SAM flags, with at least one SNP of interest in each of the forward and reverse reads using custom scripts that are deposited in the Dryad Data Repository (http://dx.doi.org/10.5061/dryad.18j31ch). When a site had the DCV-C allele, we then counted the proportion of read-pairs where the other SNP either had the DCV-C allele (non-recombinant) or the alternate allele (recombinant). As polymorphisms segregating in the founding viral populations can give a spurious signal of recombination in this analysis [38], we removed any pairs of SNPs where the ratio of reads carrying the two possible products of recombination deviated from the expected 50 : 50 ratio (binomial test, $p < 0.01$). We tested whether the fraction of recombinant reads between pairs of SNPs increased with distance between the SNPs using logistic regression, accounting for overdispersion using a quasi-binomial model. Retaining only pairs of SNPs that were represented by at least 25 read-pairs, we regressed the proportion of recombinant reads against the distance between the SNPs to estimate the fraction of recombinants between adjacent nucleotides.

(g) Drosophila C virus titre

Total RNA was extracted 3 days post-infection from pools of 10 DCV-infected flies. Five biological replicates (10 flies each) were performed per *Wolbachia* treatment, DCV isolate (single infection experiment) or DCV population (selection experiments). The extracted RNA was reverse-transcribed with Promega GoScript reverse transcriptase (Promega) and random hexamers. The amount of virus RNA was quantified with qPCR by amplifying a 135 bp region of the DCV genome with primers DCV_S (5′-GACACTGCCTTTGATTAG-3′) and DCV_AS (5′-CCCTCTGGGAACTAAATG-3′) targeting regions that are conserved among the DCV isolates used in this study. Additionally, we quantified the fly gene *actin 5C* in a separate reaction (forward: 5′-GACGAAGAAGTTGCTGCTCTGGTTG-3′; reverse: 5′-TGAGGATACCACGCTTGCTCTGC-3′; 193 bp product). The qPCR cycle was 95°C for 2 min, followed by 40 cycles of 95°C for 5 s and 60°C for 30 s. Two reactions per sample and per target gene were performed and the mean Ct value of the two technical replicates was used to calculate the relative amount of DCV RNA per fly as $2^{\Delta Ct}$, with $\Delta Ct = Ct_{\text{fly gene}} - Ct_{\text{DCV}}$.

(h) Virus-induced mortality

Flies were infected with each DCV isolate or with Ringer's solution (mock-infected controls) as above except that biological replicates consisted of vials with 20 females. Following infection, flies were transferred onto fresh food every 3 days and survival was recorded daily for 15 days post-infection.

(i) Selection experiment 2 on genetically homogeneous virus populations

In order to study virus evolution from de novo mutation, we used a similar protocol as in the selection experiment on genetically diverse virus populations. We passaged the homogeneous DCV population DCV-ref. [33] for ten passages in either *Wolbachia*-free or *Wolbachia*-infected flies (25 replicate populations in each treatment). The only difference was that male flies were used instead of females and viruses were harvested two days post-infection instead of three. Flies were initially infected in passage 1 with a virus concentration of 6.32×10^8 TCID50 ml^{-1}.

(j) Statistical analysis

All statistical analyses were done in R software v. 3.2.3 [39] and R scripts deposited in the Dryad Data Repository (http://dx.doi.org/10.5061/dryad.18j31ch). Pairwise genetic distances between virus populations of the selection experiment 1 were calculated as the probability of drawing different alleles from two different populations $P_{\text{diff}} = 1 - (p_{\text{pop1}} \times p_{\text{pop2}} + q_{\text{pop1}} \times q_{\text{pop2}})$, where p_{pop1} and p_{pop2} are the mean frequencies of reference alleles in population 1 and 2 and q_{pop1} and q_{pop2} the mean frequencies of the alternative allele. We conducted a Mantel test with 1000 permutations by randomly attributing *Wolbachia* treatments to populations. The p-value was obtained by comparing the observed mean genetic distance between treatments with the null distribution of mean genetic distances obtained by permutation.

The principal component analysis was performed with the function *dudi.pca* (R package *ade4*). For each *Wolbachia* treatment, 95% confidence ellipses were computed with the assumption of multivariate normal distribution of the data using the function *stat_ellipse* (R package *ggplot2*). The discriminant analysis of principal components was performed with the function *dapc* (R package *adegenet* [40]).

Viral titres were analysed with a linear model after $\log_{10}$ transformation of the data to meet the assumptions of normality and homoscedasticity. For the experiment measuring viral titres of evolved populations, the virus population was treated as a random effect in a linear mixed-effect model (package *LmerTest*). For the time-course analysis of DCV titre, three-parameter asymptotic exponential growth curves were fitted to the $\log_2$ viral titre using the function *nls* (R package *stats*). To test whether there was an effect of *Wolbachia* on the growth curve, we used a likelihood ratio test to compare the fit of a single curve to all the data with the fit of separate curves to the data from *Wolbachia*-infected and *Wolbachia*-free flies. To test whether there were either differences in the growth of the three viruses or there were virus-specific effects of *Wolbachia*, we compared our two-curve model (*Wolbachia* + and −) to a six-curve model (a separate curve for each of the three viruses, with and without *Wolbachia*). When plotting the curves, 95% confidence intervals were estimated by Monte Carlo simulation.

Fly survival was analysed with a Cox's proportional hazard mixed-effect model (R package *coxme*). Flies that were alive at the end of the experiment were treated as censored data. Multiple pairwise comparisons were performed with the function *glht* (R package *multcomp* [41]).

3. Results

(a) Virus populations evolve in response to *Wolbachia*

In order to test whether DCV adapts to the presence of *Wolbachia*, we passaged a genetically diverse population of

viruses through *Wolbachia*-infected or *Wolbachia*-free flies (selection experiment 1). The viral population was founded by mixing eight DCV isolates collected from both laboratory stocks and wild *D. melanogaster* from around the world [26]. The mixture initially contained equal concentrations of viral RNA from each DCV isolate and after three passages, we sequenced the polyadenylated RNA from 12 independent replicates per *Wolbachia* treatment. The mean depth of coverage of each replicate ranged from 65 to 89× (electronic supplementary material, table S2). After filtering out variants with a mean minor allele frequency below 5%, there were 167 Single Nucleotide Polymorphisms (SNPs). Among these, 161 were found among the genomes of the eight DCV isolates used to found the viral population. There were 703 SNPs among the founding DCV genomes, suggesting that approximately 77% had minor allele frequencies below 5% by the end of the selection experiment.

Parallel evolution, where the same genetic changes evolve independently in response to the same selection pressure, provides evidence of adaptation. To test for parallel evolution in our experiment, we calculated the genetic distance between all possible pairs of populations. We found that the mean genetic distance between DCV populations from the same *Wolbachia* treatment (*Wolbachia*-infected or *Wolbachia*-free flies) was less than between populations from different *Wolbachia* treatments (Mantel test, $p = 0.028$; electronic supplementary material, figure S1A). Parallel evolution of the DCV populations was also apparent in a principal component analysis (PCA) on DCV allele frequencies, where the second principle component separated DCV populations depending on whether or not they had evolved in flies infected with *Wolbachia* (figure 1*a*).

(b) *Wolbachia* alters the frequency of viral genotypes in the population

The evolutionary response to *Wolbachia* could either involve changes in the frequency of the eight founding viral genotypes or selective sweeps of specific SNPs through a recombining population of viruses. We calculated the difference in the frequency of SNPs between the populations that had evolved in flies with and without *Wolbachia* (figure 1*b*). Across the viral genome, alleles from the DCV-C isolate consistently showed higher frequencies in the presence of *Wolbachia* (figure 1*b*), while the DCV-EB, -CYG and -G alleles had lower frequencies. All of the variants that were at a substantially higher frequency in the *Wolbachia*-infected flies were present in DCV-C, and these were scattered across the genome (figure 1*b*). Alleles specific to the DCV-C isolate have increased frequency in both the *Wolbachia*-infected and *Wolbachia*-free flies, reaching mean frequencies of 77% and 67%, respectively (based on DCV-C specific alleles; electronic supplementary material, figure S1). Therefore, DCV-C was successful in all the populations, but its competitive advantage has been increased with *Wolbachia*.

These analyses suggest that there may have been limited recombination in the population, and the frequency of the founding viruses has changed in response to selection by *Wolbachia*. To confirm this, we generated longer sequence reads from the same samples and used read-pairs containing multiple SNPs to examine the rate at which DCV-C had recombined with other viral isolates. We found that there were more putative recombination events between SNPs that were further apart in the genome, as expected if there is recombination (logistic regression: $t = 4.45$, $p = 0.0001$). By estimating a per base pair recombination rate and extrapolating this to the whole genome, we estimate that 86% of DCV-C genomes will have survived intact without recombination by the end of the experiment. This supports the conclusion that we are primarily looking at changes in the frequency of the founding viruses.

As there is limited recombination between DCV isolates, we can use SNPs as markers to track changes in the frequency of different viral isolates. We first reconstructed the phylogeny of the DCV isolates using their Sanger-sequenced genomes (figure 1*c*), finding similar relationships to published analyses of restriction fragment length polymorphisms [26]. In the evolved populations, we identified isolate-specific alleles for four isolates (figure 1*d*; DCV-C, -E, -M and -T), while DCV-G and DCV-O are polymorphic since their genomes contain high numbers of ambiguous bases (electronic supplementary material, table S3). Larger numbers of SNPs were found that defined clades of viruses on the phylogeny (figure 1*c*,*d*). Comparing the frequency of these SNPs among our evolved populations confirmed that DCV-C was favoured in *Wolbachia*-infected flies, while alleles specific to the clade containing DCV-EB, -CYG and -G decreased in frequency by around 10% in the presence of *Wolbachia*. There was little difference between *Wolbachia* treatments in the frequency of isolates in the other clade on the phylogeny (DCV-M, -Z, -T and -O; figure 1*c*,*d*).

To further examine the genetic basis of differentiation between the DCV populations in *Wolbachia*-infected and *Wolbachia*-free flies, we used discriminant analysis on the principal components (DAPC). The virus populations showed a bimodal distribution which separates the *Wolbachia* treatments (electronic supplementary material, figure S2A). The genetic differentiation is driven by SNPs across the viral genome, which is consistent with there being limited recombination (electronic supplementary material, figure S2B). Alleles specific to the isolate DCV-C and the EB-CYG-G clade consistently contributed the most to the genetic differentiation between *Wolbachia* treatments (electronic supplementary material, figure S2C). This confirms that DCV-C was favoured in the presence of *Wolbachia* and that this was at the expense of viruses in the EB-CYG-G clade.

(c) The Drosophila C virus isolate favoured in *Wolbachia*-infected flies does not evade the symbiont's antiviral effects

Wolbachia may be selecting for viruses that evade its antiviral effects. In order to investigate this, we inoculated *Wolbachia*-free and *Wolbachia*-infected flies with the DCV isolate that increased the frequency in the presence of *Wolbachia* (DCV-C), one isolate that decreased in frequency (DCV-EB) and one isolate which frequency was little affected by *Wolbachia* (DCV-M). Viral titres were measured over 6 days to allow an asymptotic exponential growth curve to be fitted to the data. We found that *Wolbachia* altered the growth curve of DCV (figure 2*a*; $\chi^2 = 36.8$, d.f. = 4, $p < 0.0001$), reducing both the viral growth rate and the final viral titre. However, there was no difference in the growth curves of the three

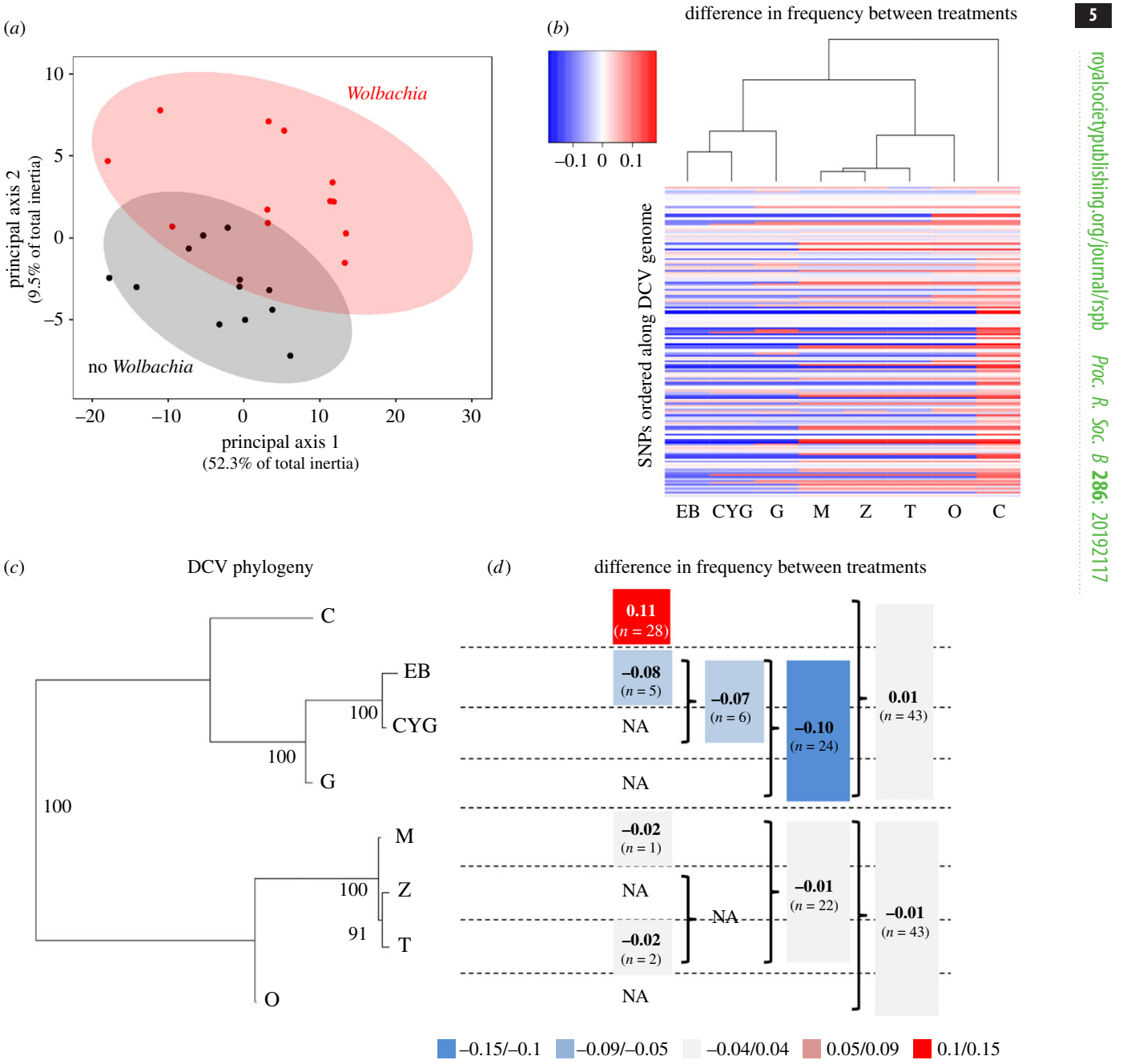

Figure 1. Effect of *Wolbachia* on virus allele frequencies when selecting on a genetically diverse viral population. A genetically diverse population of DCV was passaged through *Wolbachia*-infected or *Wolbachia*-free flies and then sequenced. (*a*) Principal components analysis on allele frequencies of SNPs, where each point is an independent virus population (biological replicate). (*b*) Differences in the frequency of SNPs along the viral genome. The sequence reads were mapped to the DCV-C reference genome, and the heatmap shows the difference in the frequency of the allele carried by a given DCV strain between *Wolbachia* treatments (frequency in *Wolbachia*-infected flies minus frequency in *Wolbachia*-free flies; red is a higher allele frequency in *Wolbachia*-infected flies). The tree was computed from the Euclidian distance computed from these differences in allele frequencies. (*c*) DCV phylogeny with bootstrap support for the nodes. (*d*) Differences in the frequency of isolate- or clade-specific variants between viral populations that had evolved in *Wolbachia*-infected and *Wolbachia*-free flies (positive numbers are more common in the presence of *Wolbachia*). The number of SNPs is shown in parentheses, and the number in bold is the mean difference in frequency across all those SNPs. NA stands for cases where no isolate- or clade-specific SNPs could be found. (Online version in colour.)

viral isolates, regardless of whether the flies were infected with *Wolbachia* (figure 2*a*; $\chi^2 = 20.6$, d.f. = 16, $p = 0.20$). Therefore, we found no evidence that *Wolbachia* has favoured viral isolates that overcome the symbiont's antiviral properties.

We also measured the ability of the DCV isolates to kill *Wolbachia*-free and *Wolbachia*-infected flies. Levels of mortality varied among virus isolates, with flies infected with DCV-CYG showing the lowest mortality, regardless of whether the flies carried *Wolbachia* (figure 2*b*; main effect virus: $\chi^2_i = 125.29$, d.f. = 10, $p < 0.0001$). *Wolbachia* had no effect on the survival of mock-infected flies (control in figure 2*b*), but in all cases increased survival of DCV-infected flies (figure 2*b*,*c*; main effect *Wolbachia*: $\chi^2_i = 132.7$, d.f. = 6, $p < 0.0001$). The magnitude of the protective effects of *Wolbachia* depended on the DCV isolate (figure 2*b*,*c*; *Wolbachia*-by-virus interaction: $\chi^2_i = 16.7$, d.f. = 2, $p = 0.005$). However, there was no association between the extent to which *Wolbachia* protected flies against the virus (figure 2*b*) and whether that virus increased the frequency in the presence of *Wolbachia* (figure 1).

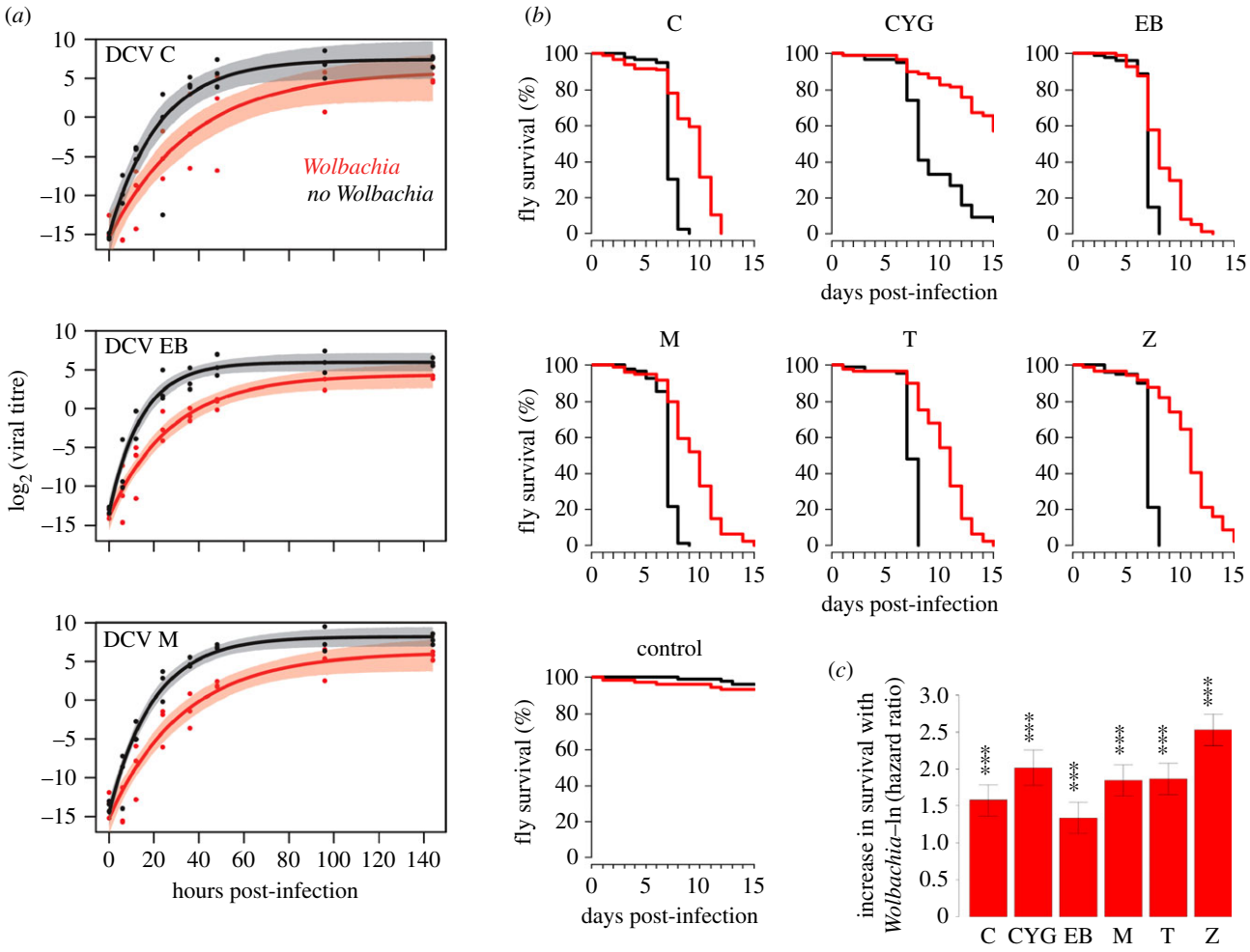

Figure 2. *Wolbachia*'s effect on viral titres and virus-induced mortality in single virus infections. (*a*) Growth curves of three DCV isolates in *Wolbachia*-free and *Wolbachia*-infected flies. The lines are asymptotic exponential curves and the shaded areas are 95% confidence intervals. (*b*) Survival curves following infection with DCV. (*c*) *Wolbachia* effect on virus-induced mortality expressed as –ln(hazard ratio) where the hazard ratio is the probability of flies dying in *Wolbachia*-infected flies relative to their *Wolbachia*-free counterpart. Error bars are standard errors and ***$p < 0.001$. (Online version in colour.)

(d) Virus populations evolved with *Wolbachia* did not adapt to counteract its antiviral effect

To test whether viruses had adapted to overcome the antiviral effects of *Wolbachia*, we randomly chose five virus populations from each of the selection treatment and infected flies with equal concentrations of viral RNA (see Methods). The viral titre 3 days post-infection was lower in *Wolbachia*-infected flies, regardless of the selection regime (*Wolbachia* effect: $\chi^2_i = 158.68$; $p < 0.0001$; electronic supplementary material, figure S3 for pairwise comparisons). Surprisingly, the effect of *Wolbachia* on viral titres was slightly greater for the viral populations that had been passaged through *Wolbachia*-infected flies (*Wolbachia*-by-selection effect: $\chi^2_i = 4.34$; $p = 0.04$; electronic supplementary material, figure S3). Therefore, viruses evolved with the symbiont were still susceptible to the inhibitory effect of *Wolbachia*.

(e) A genetically homogeneous virus population did not evolve to escape the antiviral effects of *Wolbachia*

The high mutation rates, replication rates and population sizes of many viruses mean that they can frequently evolve to overcome host resistance during selection experiments through de novo mutations. To test whether this was the case for *Wolbachia* and DCV, we serially passaged a genetically homogeneous population of the DCV-C isolate through *Wolbachia*-free and *Wolbachia*-infected flies (selection experiment 2). While DCV was maintained at high titres in *Wolbachia*-free flies, titres tended to progressively decrease in the presence of the symbiont and frequently became undetectable (figure 3*a*). After 10 passages, only 2 of 25 replicates in *Wolbachia*-infected flies had DCV titres close to those observed in the absence of the symbiont (figure 3*a*). We tested whether these two virus populations had maintained high titres by evolving to counter the antiviral effects of *Wolbachia* by infecting new flies with equal concentration of viral RNA. There was no significant difference in the effects of *Wolbachia* on these viruses compared to control populations passaged through *Wolbachia*-free flies (figure 3*b*; *Wolbachia* effect: $F_{1,44} = 113.95$; $p < 0.0001$; selection effect: $F_{1,44} = 0.72$; $p = 0.4$; *Wolbachia*-by-selection interaction: $F_{1,44} = 3.25$; $p = 0.08$).

4. Discussion

Wolbachia is able to block the replication of RNA viruses, and this is being harnessed by public health programmes to control mosquito-borne diseases [42]. The impact of such interventions on evolution of the viruses is unknown. If viruses could evolve to escape *Wolbachia*'s antiviral effects,

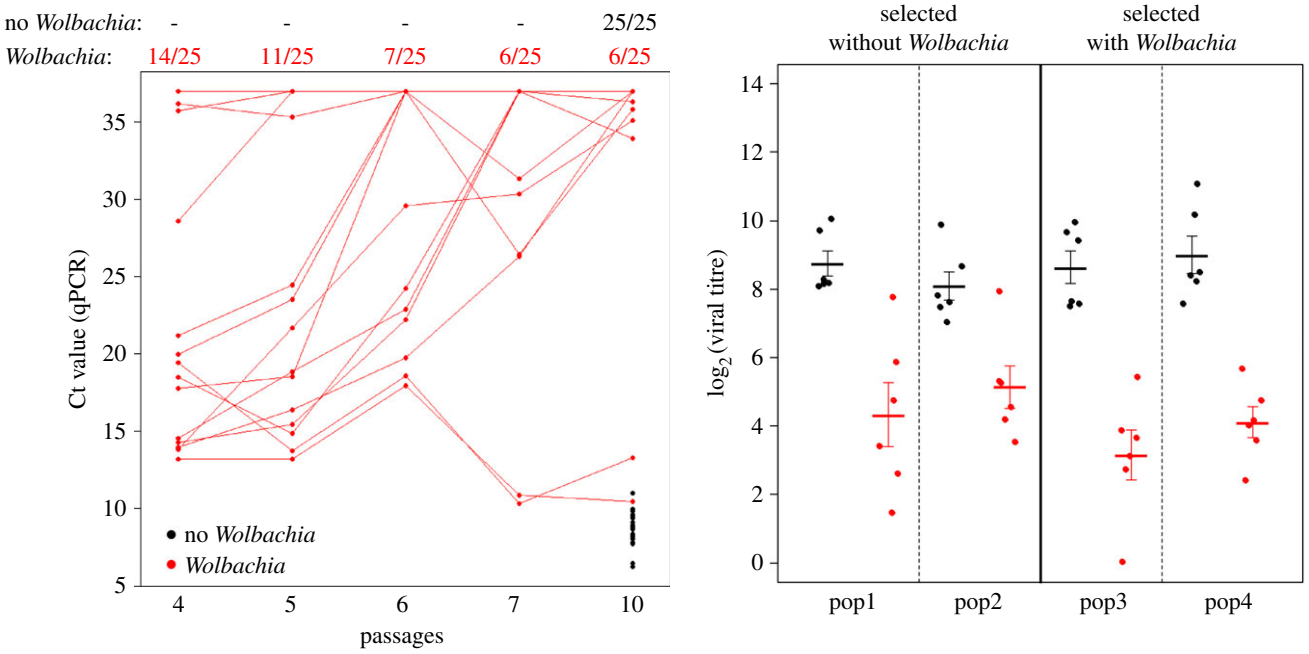

Figure 3. Presence of DCV and *Wolbachia* effect on viral titres in selection experiment 2. (*a*) Ct values obtained from qPCR reactions targeting DCV RNA in virus populations at different passages during selection. Populations passaged in *Wolbachia*-free flies were only checked for DCV infection at passage 10. Fractions at the top of the plot indicate the number of biological replicates for which DCV was detected relative to the total number of replicates ($n = 25$ in each selection treatment). (*b*) Virus titre of the DCV populations 3 days post-infection in *Wolbachia*-free (black) and -infected (red) flies. Horizontal bold lines and dots indicate mean titres and values per biological replicate, respectively. Error bars are standard errors. (Online version in colour.)

or *Wolbachia* selects for increased viral virulence in the mosquito or human host, this would have important implications for control programmes. For example, interventions that reduce the growth rate of a pathogen without clearing the infection can select for compensatory increases in replication rates that in turn increase virulence [43]. We found that *Wolbachia* can alter the evolution of a virus—when DCV evolved in the presence of *Wolbachia*, there were replicable genetic changes in viral population. This was mediated by *Wolbachia* modifying competition between virus strains within the insect. However, these changes neither allow the virus to escape *Wolbachia's* blocking effect nor alter the virus's virulence.

In our first experiment, we passaged a genetically diverse viral population through *Wolbachia*-infected flies in an attempt to select for viruses that escaped the symbiont's antiviral effects. At the genetic level, there was a clear response to selection, with the DCV-C genotype reaching higher frequencies in the presence of *Wolbachia*. However, we were unable to find any evidence that DCV-C either had higher virulence or escaped *Wolbachia's* antiviral effects. It is possible DCV-C was favoured because of *Wolbachia* altering the strength of competition between viruses. While the mechanisms of the *Wolbachia*-mediated antiviral effect remain elusive, previous studies have shown that competition between the symbiont and viruses for resources such as cholesterol might be involved [44,45]. By reducing the availability of these resources, *Wolbachia* might exacerbate fitness differences that already exist between virus isolates. Consistent with this, DCV-C increased the frequency in *Wolbachia*-free flies, but to a lesser extent than in the presence of *Wolbachia*.

In the second selection experiment, we tested virus adaptation from de novo mutation. Selection was strong, leading to the loss of most viral populations across serial passages. The few virus populations that managed to persist in the presence of the symbiont still suffered large reductions in titre in *Wolbachia*-infected flies. Therefore, we were unable to select for major-effect *Wolbachia*-escape mutations in the virus, and it is possible these viruses simply persisted due to stochastic processes. This is similar to a recent study where dengue virus populations evolved in *Wolbachia*-infected mosquito cell lines showed rapid decline and frequently went extinct [46]. Strikingly, the few dengue infections that persisted in the presence of *Wolbachia* showed an almost complete inability to replicate in both *Wolbachia*-free and *Wolbachia*-infected cells following selection. We did not observe this with DCV, but our mixed DCV populations evolved with *Wolbachia* showed slightly lower replication in *Wolbachia*-infected flies (electronic supplementary material, figure S3), i.e. they appeared to be more affected by *Wolbachia* than populations evolved without symbiont. It is possible that the presence of *Wolbachia* leads to the production of more defective viral genomes [47,48]. However, the size of this effect was modest, and whether this is a general pattern remains to be investigated.

The reason why viruses do not readily evolve to escape the antiviral effects of *Wolbachia* is unclear. However, clues come from two patterns that are recurrently observed. First, strong antiviral protection is associated with high symbiont density within cells [17,18,21]. Second, *Wolbachia* provides protection against a diverse array of distantly related RNA viruses, including viruses whose natural hosts are *Wolbachia*-infected [17,49,50]. This is reminiscent of 'quantitative' plant defences against insect herbivores [51]. These are secondary metabolites that are produced in large quantities, and protect against a broad array of herbivores by reducing the digestibility of the plant. These contrast with 'qualitative' toxin defences that are produced in low quantities [51]. While specialist herbivores frequently evolve to escape qualitative defences, quantitative defences are thought to be more

'evolution proof' [51,52]. As discussed above, one of the leading hypotheses to explain the antiviral effects of *Wolbachia* is that the symbiont competes with the virus for resources such as cholesterol [44,45]. If these resources are essential to viral replication, such a mechanism may be a form of quantitative defence that viruses cannot readily evolve to escape from. This contrasts with 'qualitative' forms of antiviral protection, such as restriction factors, changes to surface receptors or drugs, where virus escape may evolve by altering the molecular target of the antiviral effector.

The final verdict on whether *Wolbachia*-based control of vector-borne disease is more 'evolution-proof' than drugs or insecticides awaits its long-term deployment in the field. However, there are grounds for optimism. *Wolbachia* is maintained at high frequency and retains its antiviral properties years after it is released into populations [11,14]. RNA viruses, including DCV, frequently show considerable responses to selection in short-term laboratory experiments [33,53]. Therefore, our results, together with similar work in cell culture, demonstrate that viruses do not readily evolve to escape *Wolbachia's* effects [14]. Nonetheless, in the wild, viral population sizes are larger and selection will act for longer periods of time, so our results may not reflect the outcome of this interaction in nature. However, *Wolbachia* effectively protects against viruses that will probably have experienced many decades of selection because their natural hosts are *Wolbachia*-infected, suggesting that our experiments may reflect the outcome of evolution in nature [17,49,50] (although it is unclear whether the symbiont is an important selection pressure in natural populations of *D. melanogaster* [54,55]). Long-term monitoring of field populations will be essential to test whether this reflects fundamental biological constraints that prevent viruses evolving to escape from the effects of *Wolbachia*.

Data accessibility. Data and R scripts are available at the Dryad Digital Repository: https://doi.org/10.5061/dryad.18j31ch [56] and Sequence Read Archive (accession number PRJEB21984) [57].

Authors' contributions. J.M., G.B.-M., S.C.L.S., J.P.D. and B.L. carried out the experimental work. J.M., R.A., G.B.-M. and F.M.J. analysed the data. J.M. and F.M.J. designed the study and drafted the manuscript. All authors gave final approval for publication.

Competing interests. The authors have no competing interests.

Funding. This study was funded by the Wellcome Trust grant no. WT094664MA (http://www.wellcome.ac.uk/) and the European Research Council (ERC) grant no. 281668 DrosophilaInfection. G.B.-M. was supported by SENACYT-IFARHU.

References

1. Weinert LA, Araujo-Jnr EV, Ahmed MZ, Welch JJ. 2015 The incidence of bacterial endosymbionts in terrestrial arthropods. *Proc. R. Soc. B* **282**, 20150249. (doi:10.1098/rspb.2015.0249)
2. Kriesner P, Hoffmann AA, Lee SF, Turelli M, Weeks AR. 2013 Rapid sequential spread of two *Wolbachia* variants in *Drosophila simulans*. *PLoS Pathog.* **9**, e1003607. (doi:10.1371/journal.ppat.1003607)
3. Turelli M, Hoffmann A. 1991 Rapid spread of an inherited incompatibility factor in California *Drosophila*. *Nature* **353**, 440–442. (doi:10.1038/353440a0)
4. Hoffmann AA *et al.* 2011 Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. *Nature* **476**, 454–457. (doi:10.1038/nature10356)
5. Teixeira L, Ferreira A, Ashburner M. 2008 The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biol.* **6**, 2753–2763. (doi:10.1371/journal.pbio.1000002)
6. Moreira LA *et al.* 2009 A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, Chikungunya, and Plasmodium. *Cell* **139**, 1268–1278. (doi:10.1016/j.cell.2009.11.042)
7. Rainey SM *et al.* 2016 *Wolbachia* blocks viral genome replication early in infection without a transcriptional response by the endosymbiont or host small RNA pathways. *PLoS Pathog.* **12**, e1005536. (doi:10.1371/journal.ppat.1005536)
8. Dutra HLC, Rocha MN, Dias FBS, Mansur SB, Caragata EP, Moreira LA. 2016 *Wolbachia* blocks currently circulating Zika virus isolates in Brazilian *Aedes aegypti* mosquitoes. *Cell Host Microbe.* **19**, 771–774. (doi:10.1016/j.chom.2016.04.021)
9. Benelli G, Jeffries CL, Walker T. 2016 Biological control of mosquito vectors: past, present, and future. *Insects* **7**, 52. (doi:10.3390/insects7040052)
10. Schmidt TL *et al.* 2017 Local introduction and heterogeneous spatial spread of dengue-suppressing *Wolbachia* through an urban population of *Aedes aegypti*. *PLoS Biol.* **15**, 1–28. (doi:10.1371/journal.pbio.2001894)
11. Hoffmann AA, Iturbe-Ormaetxe I, Callahan AG, Phillips BL, Billington K, Axford JK, Montgomery B, Turley AP, O'neill SL. 2014 Stability of the wMel *Wolbachia* infection following invasion into *Aedes aegypti* populations. *PLoS Negl. Trop. Dis.* **8**, e3115. (doi:10.1371/journal.pntd.0003115)
12. Dutra HLC *et al.* 2015 From lab to field: the influence of urban landscapes on the invasive potential of *Wolbachia* in Brazilian *Aedes aegypti* mosquitoes. *PLoS Negl. Trop. Dis.* **9**, 1–22. (doi:10.1371/journal.pntd.0003689
13. Nguyen TH *et al.* 2015 Field evaluation of the establishment potential of wMelPop *Wolbachia* in Australia and Vietnam for dengue control. *Parasit Vectors.* **8**, 563. (doi:10.1186/s13071-015-1174-x)
14. Frentiu FD, Zakir T, Walker T, Popovici J, Pyke AT, van den Hurk A, Mcgraw EA, O'neill SL. 2014 Limited dengue virus replication in field-collected *Aedes aegypti* mosquitoes infected with *Wolbachia*. *PLoS Negl. Trop. Dis.* **8**, e2688. (doi:10.1371/journal.pntd.0002688)
15. Bull JJ, Turelli M. 2013 *Wolbachia* versus dengue: evolutionary forecasts. *Evol. Med. Public Heal.* **2013**, 197–201. (doi:10.1093/emph/eot018)
16. Vavre F, Charlat S. 2012 Making (good) use of *Wolbachia*: what the models say. *Curr. Opin. Microbiol.* **15**, 263–268. (doi:10.1016/j.mib.2012.03.005)
17. Martinez J, Longdon B, Bauer S, Chan Y-S, Miller WJ, Bourtzis K, Teixeira L, Jiggins FM. 2014 Symbionts commonly provide broad spectrum resistance to viruses in insects: a comparative analysis of *Wolbachia* strains. *PLoS Pathog.* **10**, e1004369. (doi:10.1371/journal.ppat.1004369)
18. Osborne SE, Leong YS, O'Neill SL, Johnson KN. 2009 Variation in antiviral protection mediated by different *Wolbachia* strains in *Drosophila simulans*. *PLoS Pathog.* **5**, e1000656. (doi:10.1371/journal.ppat.1000656)
19. Suh E, Dobson SL. 2013 Reduced competitiveness of *Wolbachia* infected *Aedes aegypti* larvae in intra- and inter-specific immature interactions. *J. Invertebr. Pathol.* **114**, 173–177. (doi:10.1016/j.jip.2013.08.001)
20. Martinez J, Ok S, Smith S, Snoeck K, Day JP, Jiggins FM. 2015 Should symbionts be nice or selfish? Antiviral effects of *Wolbachia* are costly but reproductive parasitism is not. *PLoS Pathog.* **11**, e1005021. (doi:10.1371/journal.ppat.1005021)
21. Chrostek E, Marialva MSP, Esteves SS, Weinert LA, Martinez J, Jiggins FM, Teixeira L. 2013 *Wolbachia* variants induce differential protection to viruses in *Drosophila melanogaster*: a phenotypic and phylogenomic analysis. *PLoS Genet.* **9**, e1003896. (doi:10.1371/journal.pgen.1003896)
22. Chrostek E, Teixeira L. 2015 Mutualism breakdown by amplification of *Wolbachia* genes. *PLoS Biol.* **13**, e1002065. (doi:10.1371/journal.pbio.1002065)
23. Ferguson NM *et al.* 2015 Modeling the impact on virus transmission of *Wolbachia*-mediated blocking of dengue virus infection of *Aedes aegypti*. *Sci.*

Transl. Med. **7**, 279ra37. (doi:10.1126/scitranslmed.3010370)

24. Plus N, Croizier G, Jousset F, David J. 1975 Picornaviruses of laboratory and wild *Drosophila melanogaster*: geographical distribution and serotypic composition. *Ann. Microbiol. (Paris)* **126**, 107–117.
25. Arnold PA, Johnson KN, White CR. 2013 Physiological and metabolic consequences of viral infection in *Drosophila melanogaster*. *J. Exp. Biol.* **216**(Pt 17), 3350–3357. (doi:10.1242/jeb.088138)
26. Johnson KN, Christian PD. 1999 Molecular characterization of Drosophila C virus isolates. *J. Invertebr. Pathol.* **73**, 248–254. (doi:10.1006/jipa.1998.4830)
27. Sullivan W, Ashburner M, Hawley R. 2000 *Drosophila protocols*. New York, NY: Cold Spring Harbor Laboratory Press.
28. Reed L, Muench H. 1938 A simple method of estimating fifty per cent endpoints. *Am. J. Hyg.* **27**, 493–497. (doi:10.1093/oxfordjournals.aje.a1184(08))
29. Hall T. 1999 BioEdit: a user-friendly biological sequences alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98.
30. Martinez J, Tolosana I, Ok S, Smith S, Snoeck K, Day JP, Jiggins FM. 2017 Symbiont strain is the main determinant of variation in *Wolbachia*-mediated protection against viruses across *Drosophila* species. *Mol. Ecol.* **26**, 4072–4084. (doi:10.1111/mec.14164)
31. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170)
32. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760. (doi:10.1093/bioinformatics/btp324)
33. Longdon B, Day JP, Alves JM, Smith SCL, Houslay TM, McGonigle JE, Tagliaferri L, Jiggins FM. 2018 Host shifts result in parallel genetic changes when viruses evolve in closely related species. *PLoS Pathog.* **14**, e1006951. (doi:10.1371/journal.ppat.1006951)
34. DePristo MA *et al.* 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491. (doi:10.1038/ng.806)
35. Li H *et al.* 2009 The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078–2079.
36. Kofler R, Orozco-terWengel P, de Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C. 2011 Popoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE* **6**, e15925. (doi:10.1371/journal.pone.0015925)
37. Lindenbaum P. 2015 JVarkit: Java-based utilities for bioinformatics. See https://github.com/lindenb/jvarkit.
38. Hudson RR, Kaplan NL. 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147–164.
39. R Core Team. 2013 *R: a language and environment for statistical computing* (ed. PF Pimenta). Vienna, Austria: R Foundation for Statistical Computing.
40. Jombart T, Devillard S, Balloux F. 2010 Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94. (doi:10.1186/1471-2156-11-94)
41. Hothorn T, Bretz F, Westfall P. 2008 Simultaneous inference in general parametric models. *Biom. J.* **50**, 346–363. (doi:10.1002/bimj.200810425)
42. Neill SLO *et al.* 2019 Scaled deployment of Wolbachia to protect the community from dengue and other Aedes transmitted arboviruses. *Gates Open Res.* **2**, 36. (doi:10.12688/gatesopenres.12844.3)
43. Read AF, Baigent SJ, Powers C, Kgosana LB, Blackwell L, Smith LP, Kennedy DA, Walkden-Brown SW, Nair VK. 2015 Imperfect vaccination can enhance the transmission of highly virulent pathogens. *PLoS Biol.* **13**, e1002198. (doi:10.1371/journal.pbio.1002198)
44. Caragata EP, Rancès E, Hedges LM, Gofton AW, Johnson KN, O'Neill SL, Mcgraw EA. 2013 Dietary cholesterol modulates pathogen blocking by *Wolbachia*. *PLoS Pathog.* **9**, e1003459. (doi:10.1371/journal.ppat.1003459)
45. Geoghegan V *et al.* 2017 Perturbed cholesterol and vesicular trafficking associated with dengue blocking in *Wolbachia*-infected *Aedes aegypti* cells. *Nat. Commun.* **8**, 526. (doi:10.1038/s41467-017-00610-8)
46. Koh C, Audsley MD, Di Giallonardo F, Kerton EJ, Young PR, Holmes EC, McGraw EA. 2019 Sustained *Wolbachia*-mediated blocking of dengue virus isolates following serial passage in *Aedes aegypti* cell culture. *Virus Evol.* **5**, 1–9. (doi:10.1093/ve/vez012)
47. Juarez-Martinez AB, Vega-Almeida TO, Salas-Benito M, Garcia-Espitia M, De Nova-Ocampo M, Del Angel RM, Salas-Benito JS. 2013 Detection and sequencing of defective viral genomes in C6/36 cells persistently infected with dengue virus 2. *Arch. Virol.* **158**, 583–599. (doi:10.1007/s00705-012-1525-2)
48. Aaskov J, Buzacott K, Thu HM, Lowry K, Holmes EC. 2006 Long-term transmission of defective RNA viruses in humans and *Aedes* mosquitoes. *Science* **311**, 236–238. (doi:10.1126/science.1115030)
49. Mousson L, Zouache K, Arias-Goeta C, Raquin V, Mavingui P, Failloux A-B. 2012 The native *Wolbachia* symbionts limit transmission of dengue virus in *Aedes albopictus*. *PLoS Negl. Trop. Dis.* **6**, e1989. (doi:10.1371/journal.pntd.0001989)
50. Glaser RL, Meola MA. 2010 The native *Wolbachia* endosymbionts of *Drosophila melanogaster* and *Culex quinquefasciatus* increase host resistance to West Nile virus infection. *PLoS ONE* **5**, e11977. (doi:10.1371/journal.pone.0011977)
51. Smilanich AM, Fincher RM, Dyer LA. 2016 Does plant apparency matter? Thirty years of data provide limited support but reveal clear patterns of the effects of plant chemistry on herbivores. *New Phytol.* **210**, 1044–1057. (doi:10.1111/nph.13875)
52. Rhoades DF, Cates RG. 1976 Toward a general theory of plant antiherbivore chemistry. In *Biochemical interaction between plants and insects* (eds JW Wallace, RL Mansell), pp. 168–213. Boston, MA: Springer US.
53. Coffey LL, Vasilakis N, Brault AC, Powers AM, Tripet F, Weaver SC. 2008 Arbovirus evolution in vivo is constrained by host alternation. *Proc. Natl Acad. Sci. USA* **105**, 6970–6975. (doi:10.1073/pnas.0712130105)
54. Shi M, White VL, Schlub T, Eden J-S, Hoffmann AA, Holmes EC. 2018 No detectable effect of *Wolbachia w* Mel on the prevalence and abundance of the RNA virome of *Drosophila melanogaster*. *Proc. R. Soc. B* **285**, 20181165. (doi:10.1098/rspb.2018.1165)
55. Webster CL *et al.* 2015 The discovery, distribution, and evolution of viruses associated with *Drosophila melanogaster*. *PLoS Biol.* **13**, e1002210. (doi:10.1371/journal.pbio.1002210)
56. Martinez J, Bruner-Montero G, Arunkumar R, Smith SCL, Day JP, Longdon B, Jiggins FM. 2019 Data from: Virus evolution in *Wolbachia*-infected *Drosophila*. Dryad Digital Repository. (doi:10.5061/dryad.18j31ch)
57. Martinez J, Bruner-Montero G, Arunkumar R, Smith SCL, Day JP, Longdon B, Jiggins FM. 2019 Data from: Virus evolution in *Wolbachia*-infected *Drosophila*. Sequence Read Archive. See www.ncbi.nlm.nih.gov/bioproject/PRJEB21984)