Genetic stability of *Aedes aegypti* populations following invasion by *w*Mel *Wolbachia*

Meng-Jia Lau[1]*, Tom Schmidt[1]*, Qiong Yang[1], Jessica Chung[1,2], Lucien Sankey[1], Perran A. Ross[1], Ary A. Hoffmann[1]

[1] Pest and Environmental Adaptation Research Group, Bio21 Institute and the School of BioSciences, The University of Melbourne, Parkville, Victoria, Australia

[2] Melbourne Bioinformatics, University of Melbourne, Parkville, VIC, Australia

* Corresponding author

Meng-Jia Lau: mengjial2@student.unimelb.edu.au

Tom Schmidt: tom.schmidt@unimelb.edu.au

Ary Hoffmann: ary@unimelb.edu.au

Abstract

Background

Wolbachia *w*Mel is the most used strain in mosquito rear and release strategies that aim to inhibit the transmission of arboviruses such as dengue, Zika, Chikungunya and yellow fever. However, the long-term establishment of *w*Mel in natural populations of the dengue mosquito *Aedes aegypti* raises concerns that interactions between *Wolbachia* *w*Mel and *Ae. aegypti* may lead to changes in the host genome, which could affect useful attributes of *Wolbachia* that allow it to invade and suppress disease transmission.

Results

We applied an evolve-and-resequence approach to study genome-wide genetic changes in *Ae. aegypti* from the Cairns region, Australia, where *Wolbachia* *w*Mel was first introduced more than 10 years ago. Mosquito samples were collected at three different time points in Gordonvale, Australia, covering the phase before (2010) and after (2013 and 2018) *Wolbachia* releases. An additional three locations where *Wolbachia* replacement happened at different times across the last decade were also sampled in 2018. We found that the genomes of mosquito populations mostly remained stable after *Wolbachia* release, with population differences tending to reflect the geographic location of the populations rather than *Wolbachia* infection status. However, outlier analysis suggests that *Wolbachia* may have had an influence on some genes related to immune response, development, recognition and behavior.

Conclusions

Aedes aegypti populations remained geographically distinct after *Wolbachia* releases in North Australia despite their *Wolbachia* infection status. At some specific genomic

loci, we found signs of selection associated with *Wolbachia*, suggesting potential evolutionary impacts can happen in the future and further monitoring is warranted.

Keywords

Wolbachia, *Aedes aegypti*, evolve-and-resequence, population replacement, Bayesian analysis,

Background

Wolbachia are bacteria that live inside the cells of many insects and induce a number of important phenotypic effects on their hosts that can be harnessed for pest and disease control. *Wolbachia*-infected *Aedes aegypti* mosquitoes have now been released in multiple locations of the world [1-3] to help reduce the transmission of arboviruses such as dengue, Zika, Chikungunya and yellow fever [4-6]. *Wolbachia* *w*Mel, which was transferred artificially from *Drosophila melanogaster* into *Ae. aegypti* [6], was first released in Gordonvale and Yorkeys Knob, Queensland, Australia, around a decade ago [7] where it invaded the local population through cytoplasmic incompatibility (CI). CI results in uninfected females less likely to produce viable offspring if they mate with infected males. In contrast, infected females produce viable offspring when they mate with uninfected males or males infected by the same *Wolbachia* strain, and these offspring are infected [8]. This allows *Wolbachia* to invade and be self-sustained in a population but may also increase population divergence because it can reduce the "effective migration rate" [9, 10] between infected and uninfected populations. *Wolbachia* can also impact mitochondrial DNA (mtDNA) variation through indirect linkage disequilibrium [11-13].

With the success of *Wolbachia* in suppressing dengue following invasion [1, 3], many studies have now focused on the sustainability of this approach beyond the initial spread, such as the maintenance of high infection levels [14], fitness costs [15, 16] and evolutionary adaptation [17, 18]. The potential evolutionary changes in *Wolbachia* *w*Mel-infected *Ae. aegypti* as well as in the bacterial genome itself following releases in the field have raised concerns about the long-term effectiveness of the strategy. The genetic background of the mosquito host can affect the capacity of *Wolbachia* to invade populations and suppress arboviruses [2, 19, 20]. *Aedes aegypti* has a short generation interval of ~1 month, and so if the introduction of *Wolbachia* triggers an evolutionary process in the mosquito genome this could be observable within a few years after invasion. Evolutionary changes in response to natural *Wolbachia* infections have previously been noted and appear to involve both the *Wolbachia* and host genomes [21, 22], affecting the population dynamics of *Wolbachia* infections. Adaptations can occur to counter any negative fitness effects of *Wolbachia*, as documented in *Drosophila* [21, 23], and negative fitness effects are particularly evident in novel infections transfected into new hosts [24].

In the past decade, the *w*Mel infection itself has not evolved in terms of either sequence composition or density since establishment in *Ae. aegypti* in northern Queensland, Australia [25]. Phenotypic comparisons also suggest limited changes in host fitness costs and CI since population replacement in this region [14, 18], although the number of fitness-related traits scored so far has been limited. Blockage of virus transmission also appears stable to date [26, 27], and may persist through ongoing selection favoring high viral blocking in *Ae. aegypti* populations [19].

There are no published studies investigating evolutionary changes in wild host populations at the genomic level following a *Wolbachia* release. Any putative changes may guide further phenotypic comparisons based on the types of candidate genes identified. However, there are challenges in characterizing genome changes in *Ae. aegypti* after *Wolbachia w*Mel infection. First, the genome of *Ae. aegypti* contains a large proportion (47%) of transposable elements (TEs), which result in a large genome size (1.38Gb) compared to other mosquitoes [28-30]. TEs might also enhance rates of evolution, given that they are involved with gene regulation, and increase genome plasticity [31]. Moreover, the involvement of other environmental factors in field-collected samples cannot be neglected, impacts of gene flow following the activity of *Wolbachia* release will also increase the difficulty of outlier analyses. Finally, compared to model organisms, the genome of *Ae. aegypti* is still relatively poorly annotated, with only 256 proteins (<1%) reviewed in the Swiss-Prot database (https://www.uniprot.org/taxonomy/7159).

In this study, we analyse pooled whole genome sequencing (WGS) data of mosquitoes from Gordonvale, Australia, where releases first took place, covering three different time points from the pre- and post-release phase. As a comparison, we also sequenced samples in Edge Hill, Redlynch and Yorkeys Knob, where *Wolbachia* replacement happened at different times across the last decade. We combined analyses of spatial variation in *Wolbachia* infection status and of genetic diversity in the mosquitoes to reveal the potential impact of *Wolbachia w*Mel on the genome of its host *Ae. aegypti*.

Results

Genetic variation in *Aedes aegypti* populations

Aedes aegypti were collected from four sites around Cairns, Australia, at different times pre- and post-*Wolbachia* release (Fig. 1, Table 1). We investigated patterns of genetic variation within populations using PoPoolation v. 1.2.2 [32] with the genomic annotation file from the reference AaegL5.0. In Tajima's pi (nucleotide diversity π) analysis (Fig. 1), differences we found among the populations depended on chromosomal location. We analysed the genome-wide nucleotide diversity in 10 kbp non-overlapping windows. The nucleotide diversities were different among populations, while in areas near the center of each chromosome, diversities were low and similar for all populations. This pattern has been shown in previous sequencing studies [30, 33]. Overall, nucleotide diversity was highest on chromosome 1, which

contains the sex determining locus and contains relatively lower gene densities and TEs but higher satellites compared to chromosomes 2 and 3 [34].

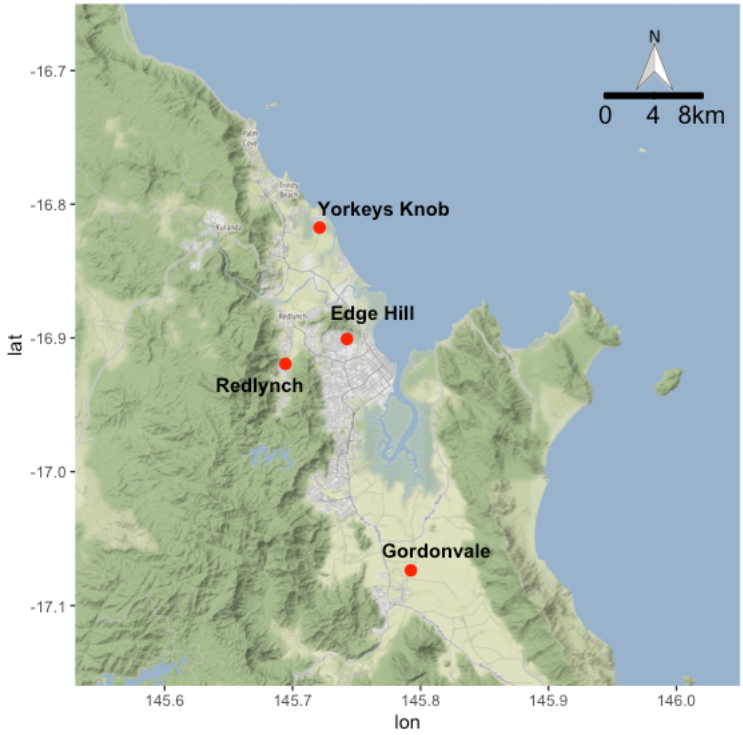

Fig. 1. Locations of sampled *Aedes aegypti* populations. Samples from Gordonvale were collected in 2010, 2013 and 2018, while samples from other locations were collected in 2018. Axes show longitude (lon) and latitude (lat).

Table 1. Summary of *Aedes aegypti* collections and designations of samples used in comparisons

Population name	Collecting year	Location	*Wolbachia* infection (Y/N)	Year of population replacement	Sample size
GV10	2010	Gordonvale	N	2011	56
GV13	2013	Gordonvale	Y	2011	51
GV18	2018	Gordonvale	Y	2011	52
EH	2018	Edge Hill	Y	2013	52
RL	2018	Redlynch	N	2019	52
YK	2018	Yorkeys Knob	Y	2011	52

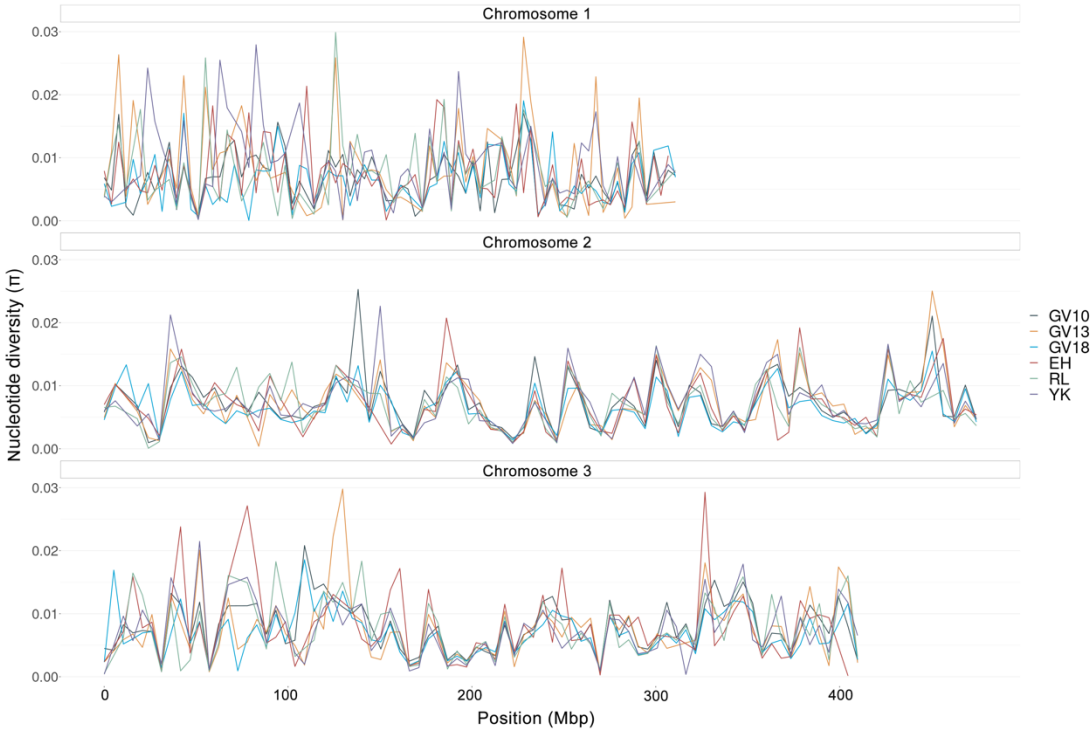

Fig. 2. LOESS-smoothed curves of genome-wide nucleotide diversity (π). Six populations of *Ae. aegypti* measured in 10 kbp non-overlapping windows. GV10 and GV13 represent samples collected in 2010 and 2013 from Gordonvale; GV18, EH, YK and RL represent samples collected in 2018 from Gordonvale, Yorkeys Knob, Edge Hill and Redlynch respectively.

In Tajima's D analysis, we found that the density distributions of values were similar between populations, except for Gordonvale pre-release (GV10) which shows a high proportion of negative values (Fig. 3). This pattern is more obvious at the gene level (Fig. 3b) than at the genome level measured in 10 kbp non-overlapping windows (Fig. 3a). The four 2018 populations converge regardless of their *Wolbachia* status or time since *Wolbachia* was invaded. This suggests that the pattern reflects a difference in GV10 before release rather than an effect of *Wolbachia per se*.

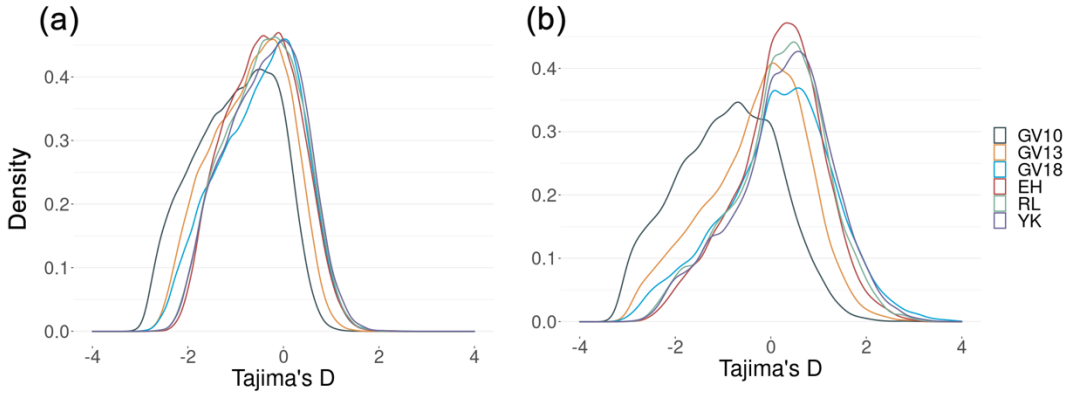

Fig. 3. Density distributions of Tajima's D values. Tajima's D measured at (a) the genome level measured in 10 kbp non-overlapping windows and at (b) the gene level. GV10 and GV13 represent samples collected in 2010 and 2013 from Gordonvale; GV18, EH, YK and RL represent samples collected in 2018 from Gordonvale, Yorkeys Knob, Edge Hill and Redlynch respectively.

Geographic segregation of *Aedes aegypti* populations

We obtained a genetic distance matrix from the average of pairwise Fst (Fixation index) values through 100 kbp non-overlapping windows (Additional file 1), with temporally-separated GV samples tending to have lower Fst values than comparisons across geographically separated samples. In the analysis of isolation by distance, with only four 2018 populations included, the genetic distance had a weak but non-significant correlation with log transformed geographical distance (road distance) in a Mantel test (r=0.66, p=0.12).

There were 461,067 SNPs left after filtering with minimal depth of 50 in all populations and an average MAF (minor allele frequency) more than 0.1. We did a principal components analysis based on these SNPs, with PC1 and PC2 accounting for 24.2% and 22.9% of the variance respectively (Fig. 4a). The three temporally-seperated GV samples fell out together, but GV10 was closer to GV18 than to GV13. We also did a PCA analysis on pairwise Fst differences in 100 kbp non-overlapping windows across the genome (Fig. 4b, d) or at the gene level (Fig. 4c, e). When testing the similarity of infected or uninfected populations (Fig. 4d, e), we found little evidence for any clustering of populations related to *Wolbachia* infection status either across the genome or at the gene level. On the other hand, these populations were separated geographically (Fig. 4b, c); the pairwise comparisons of GV samples and samples located away from GV tended to cluster when comparing across the genome, and were more separated at the gene level.

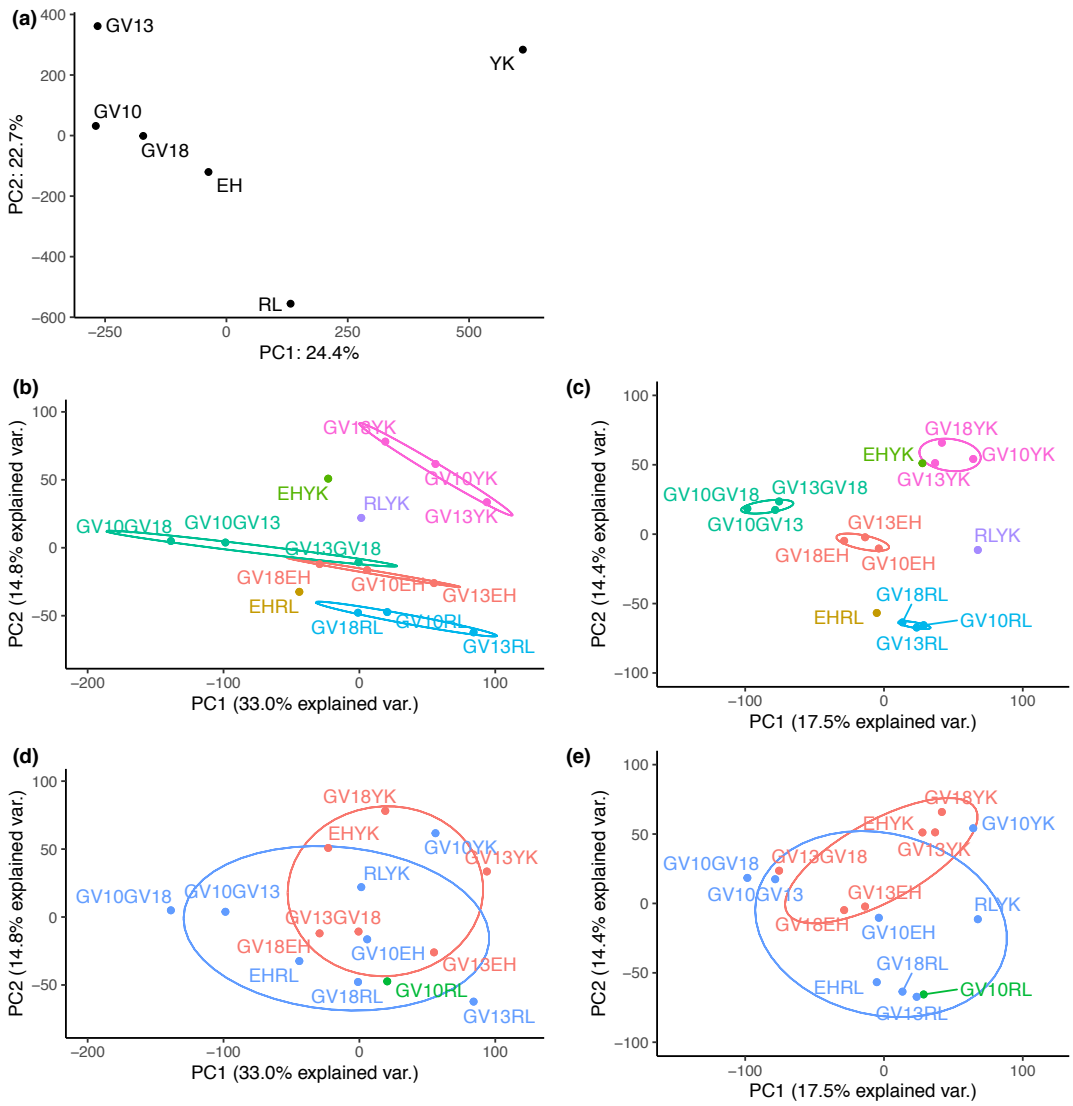

Fig. 4. Principal components analysis based on MAF or pairwise Fst. PCA plots of (a) allele frequency of *Aedes aegypti* samples (MAF > 0.1, minimal coverage > 50); (b, d) pairwise Fst throughout the genome with 100 kbp non-overlapping windows, and (c, e) pairwise Fst within genes. Colors in (b, c) represent comparisons between GV samples and a sample from a different location. The blue color in (d, e) represents comparisons between *Wolbachia*-infected and uninfected samples, while red represents comparisons within *Wolbachia*-infected samples and green represents the comparison within uninfected samples.

Bayesian models to identify outliers potentially associated with *Wolbachia*

To investigate potential selection associated with *Wolbachia*, we introduced 461,067 SNPs from the above filtering process and used two Bayesian models from BayPass v. 2.2 [35] for *Wolbachia*-related outlier analysis.

In the covariate model of BayPass, with *Wolbachia* infection status in the comparison, we found 2415 SNPs showing a "substantial" signature of selection with an average

BF* > 5 (Bayes Factor in dB units ($BF^* = 10 \times \log10(BF)$), and 391 showing "strongly-selected" signature of selection with an average BF* > 10 (Fig. 5).

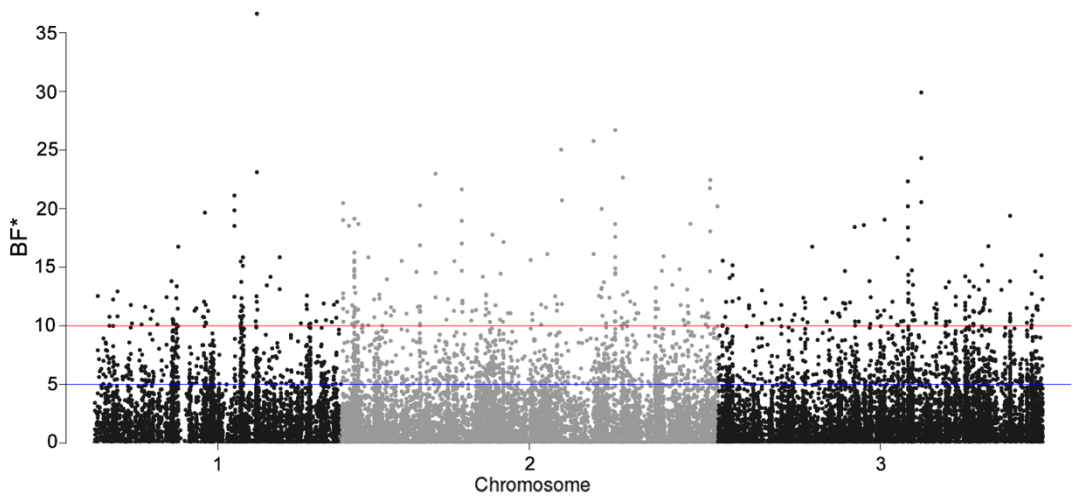

Fig. 5. Manhattan plot of SNPs in covariate model representing *Wolbachia* infection status. Only SNPs with a positive Bayes Factor are shown. The horizontal blue lines represent the cut-off for SNPs with BF* > 5 (2415 outliers) and the horizontal red lines represent the cut-off for SNPs with BF* > 10 (391 outliers).

The introduction of linear relationships in the covariate model, however, can have high false positives from sampling noise in particular when large environmental effects are involved and small number of populations are compared. We therefore used a second model to calculate X^TX, which was analogous to Fst [35, 36], then combined these two models to identify outliers potentially associated with *Wolbachia* infection.

We found that 950 (prior probability: 0.74 in each Mbp) of the "substantial" SNPs fell into the intersection of top 10% X^TX values in the comparisons between GV10 and GV13 and between GV10 and GV18, including 229 SNPs distributed on chromosome 1, 390 on chromosome 2 and 330 on chromosome 3, while one was found on an assembly scaffold NW_018735222.1. A proportion of these SNPs were concentrated in some regions, with posterior probability at least five times greater than prior probability (> 3.7 in each Mbp), suggesting selection on *Wolbachia* (Table 2). These SNPs were considered as "substantial" outliers associated with the *Wolbachia* infection.

We also used a stricter criterion: average BF*>=10 and X^TX values falling into the intersection of 95% threshold in the GV10, GV13 comparison and the GV10, GV18 comparison. We then found 113 SNPs that were highly associated with *Wolbachia* infection and were considered as "strong" outliers.

Table 2. Selected regions impacted by *Wolbachia* infection status.

Chromosome	Position	Size (Mbp)	Number of SNPs
1	96900000-105500000	8.6	39
1	138000000-140500000	2.5	13
1	148500000-149500000	1.0	11
1	183000000-188300000	5.3	40
1	203100000-206200000	3.1	22
2	15600000-16800000	1.2	27
2	20700000-21200000	0.5	11
2	98800000-99300000	0.5	18
2	345200000-345400000	0.2	19
2	396100000-396500000	0.4	15
3	239700000-249400000	9.7	31
3	311900000-316000000	4.1	25
3	368400000-368600000	0.2	14

Pathway analysis and gene ontology enrichment analysis

The "substantial" outliers were distributed within 187 genes (Additional file 2). There were 1436 genes potentially involved when a 100-kbp region around the outliers was taken into account (Additional file 3). We performed a pathway analysis of these 1436 genes through the KEGG database and found eight pathways significantly involved (Table 3, Additional file 4), including pathways involved with development and immune response (MAPK signalling pathway and Toll and Imd signaling pathway). These genes were also BLAST searched for homologues in *Drosophila melanogaster*; after filtering out the low-quality matches, we obtained 577 homologues proteins (Additional file 5). In the gene ontology enrichment (GO) analysis, 108 gene sets were significantly enriched in biological process (Additional file 6), 42 in cellular components (Additional file 7) and 22 in molecular factors (Additional file 8). In general, we found that *Wolbachia w*Mel may modulate genes with diverse functions such as cell development, interaction, cellular response, cellular transport, neurogenesis, lipid and glucide metabolization, behavior and immune response.

Table 3. Significantly enriched pathways in KEGG database potentially involved with *Wolbachia* infection.

Pathway	Description	Gene Ratio	Bg Ratio	P value
aag00232	Caffeine metabolism	6/296	12/3281	<0.001
aag00981	Insect hormone biosynthesis	9/296	40/3281	0.008
aag00052	Galactose metabolism	7/296	30/3281	0.015
aag04341	Hedgehog signaling pathway - fly	6/296	25/3281	0.021
aag04013	MAPK signaling pathway - fly	14/296	92/3281	0.034
aag00230	Purine metabolism	14/296	96/3281	0.047

aag00270	Cysteine and methionine metabolism	6/296	30/3281	0.048
aag04624	Toll and Imd signaling pathway	9/296	54/3281	0.050

The "strong" outliers were distributed across 31 genes (Table 4), including some interesting ones, such as cytochrome P450 gene 5564751, which was found enriched in response to dengue virus infection in refractory mosquitoes [37], and associated with insecticide resistance [38, 39]. Carbonic anhydrase gene 5565700 has the function of balancing pH in mosquito midgut [40, 41]. Ecdysone protein E75, encoded by gene 5569135, and lipophorin, encoded by gene 5572681, were highly expressed in females after blood feeding, potentially involved in regulation of oogenesis and vitellogenesis [42, 43]. Charged multivesicular body protein, encoded by gene 5573292, was associated with endosome formation and can influence mosquito immune response [44].

Table 4. "strong" outliers associated with *Wolbachia* infection are distributed in 31 genes.

GeneID	Strand	Product
5563771	-	tyrosine-protein kinase Btk29A, transcript variant X3
5564092	-	dynein heavy chain 12, axonemal
5564751	-	cytochrome P450 9e2
5564757	+	uncharacterized LOC5564757
5565513	+	glycogenin-1, transcript variant X7
5565700	-	carbonic anhydrase
5565906	+	rho guanine nucleotide exchange factor 7, transcript variant X1
5565937	+	protein serrate
5566564	+	tyrosine-protein phosphatase non-receptor type 9, transcript variant X2
5566587	-	transcription factor grauzone
5567217	+	glutactin, transcript variant X2
5567881	+	centrosomal protein of 135 kDa, transcript variant X2
5568339	+	uncharacterized LOC5568339, transcript variant X1
5568340	+	ribonuclease P protein subunit p14
5569034	+	uncharacterized protein DDB_G0271670, transcript variant X2
5569135	+	ecdysone-inducible protein E75, transcript variant X1
5570156	-	conserved oligomeric Golgi complex subunit 3
5570315	+	mediator of RNA polymerase II transcription subunit 12
5570620	-	high-affinity choline transporter 1
5570732	+	DNA replication factor Cdt1
5570937	-	protein outspread, transcript variant X5

5572681	+	apolipophorins
5573292	+	charged multivesicular body protein 6, transcript variant X1
5573627	+	kinesin-like protein costa
5575671	+	xanthine dehydrogenase
5575737	-	SPARC-related modular calcium-binding protein 2, transcript variant X2
5576373	+	mucin-5AC
5576645	-	thymus-specific serine protease
5579338	-	integrin alpha-PS1, transcript variant X3
110673979	+	mitochondrial intermembrane space import and assembly protein 40
110675820	+	vesicle-associated membrane protein 2, transcript variant X5

Discussion

We show that *Ae. aegypti* populations in Cairns remain geographically distinct following releases of *w*Mel, but also find some suggestive evidence for evolutionary changes in mosquito populations. When interpreting the results, it is important to consider the release process and target population and the fact that *Wolbachia*-induced CI can increase population divergence by reducing the migration rate across host populations when only one or both (in the case of different *Wolbachia* strains) are infected [9, 10]. The population released involved a *w*Mel transinfected strain that had been repeatedly backcrossed to a Cairns field population background, with the expectation that the released background would be 99.9% Cairns [7]. For releases in Yorkeys Knob and sites around Cairns, we did not expect the release population to differ from the background population because there is movement of mosquitoes around this area as reflected by the occasional detection of *Wolbachia*-infected mosquitoes after the release [7]. On the other hand, Gordonvale is a relatively isolated population which may have its own seasonal dynamics. Although this population is not genetically isolated based on microsatellite and EPIC markers [45], it does appear to be somewhat separated based on the SNP markers used in the current study. This may account for the pattern noted for Tajima's D where the 2010 Gordonvale population was a clear outlier.

When releasing mosquitoes and following invasion by *Wolbachia*, there is not only complete replacement of the uninfected mosquito population by *Wolbachia*-infected mosquitoes but also replacement of the mtDNA that can hitchhike along with the *Wolbachia* [12]. Also, while any linkage disequilibrium between the *Wolbachia* and nuclear DNA variants is expected to break down relatively quickly [46], new alleles may nevertheless be introduced into the population. The nuclear DNA constitution of the population might be expected to become more like the release stock for a period as released females and their offspring mate with released and resident males, although local selection should then lead to populations becoming more like the

original population. In our case, the genetic similarity between the Gordonvale populations after release and the original population prior to release might reflect local selection and ongoing introgression of the release stock with the resident population, as GV10 and GV18 are closer than GV13 in the PCA analysis. Furthermore, Yorkeys Knob and Edge Hill remain distinct from each other despite previously being invaded by the same release stock [3, 7].

Populations were more segregated at the gene level than at the genome level, which may be a consequence of a high mobility of TEs associated with ORF regions [21]. On the other hand, selection in response to local conditions and the impact of *w*Mel on *Ae. aegypti* may nevertheless influence patterns of genetic differentiation at specific loci [19]. Our Bayesian outlier analysis identified several regions in each chromosome and genes related to immune response, development, recognition and behavior that may have been under selection. These potential evolutionary impacts of *Wolbachia w*Mel on the genome of *Ae. aegypti* in the field suggest that further monitoring is warranted, although at this stage other factors unrelated to *Wolbachia* appear to have a larger impact on genomic differentiation among samples.

We found signs of selection on the Toll and Imd signalling pathway in KEGG analysis; these are important pathways in immune systems [47, 48] and viral blocking processes [48-50]. Previous transcriptomic studies of *Wolbachia* in *Ae. aegypti* showed up-regulation of these pathways in both *w*Mel and *w*MelPop-infected *Ae. aegypti* [51, 52]. Viral blocking genes are mainly distributed on chromosome 1, in addition to genes related to cytoskeleton, cell-cell adhesion and signal transduction [19]; these genes also show up in our GO analysis. Other than viral blockage, caffeine metabolism was strongly impacted, which may impact hormone metabolism and detoxification when cytochrome P450 is involved [53, 54]. We also detected enriched pathways involved with development, such as insect hormone biosynthesis and the Hedgehog signaling pathways. In GO enrichment analysis these are represented in cell growth, structure, recognition and behavior.

Conclusions

*Wolbachia w*Mel-infected *Ae. aegypti* mosquitoes have been released successfully in the field to help reduce the transmission of arboviruses, but interactions between *w*Mel and *Ae. aegypti* could result in adaptation [55, 56], altering viral blocking efficiency [19, 57], host fecundity [21] and insecticide resistance [58]. In this study, we have identified *Ae. aegypti* populations as being geographically distinct despite their *Wolbachia* infection status. However, selection associated with *Wolbachia* may still have influenced variation at some loci. This is the first time that genome evolution associated with *Wolbachia* infection has been examined in field populations where there has been a deliberate release. However, it is hard to draw conclusions about long-term impacts of *Wolbachia* on the *Aedes* genome, which may take more time to develop, and which may be different in regions where dengue is endemic, unlike in Australia. Our findings highlight the possibility that the effect of *Wolbachia*

can interact with the host genomic background, which has been shown previously based on phenotypic assays on the longevity of *w*MelPop in *Drosophila* [59].

Methods

Samples and study sites

Aedes aegypti samples were collected from four sites around Cairns (Fig. 1, Table 1). In Gordonvale, samples were collected three times: in the summer of 2010 (pre-release), as well as in 2013 and 2018 (2 – 7 years post release given that the area was stably invaded in 2011 [7]). Samples from Yorkeys Knob, Edge Hill and Redlynch were collected in 2018. Yorkeys Knob experienced *Wolbachia* invasion at the same time as Gordonvale in 2011, and Edge Hill which was invaded in 2013 [60]; Redlynch was an uninfected area when sampled in 2018. Gordonvale samples from 2010 and 2013 were collected by BG-Sentinel traps (Biogents, Regensburg, Germany) while 2018 samples were collected by ovitraps, taking care to sample only 1-2 larvae per ovitrap to reduce the likelihood of siblings being sampled [61]. Samples from Gordonvale 2010 (GV10) and 2013 (GV13) were stored in 100% ethanol at the adult stage while samples collected from 2018 (GV18, EH, YK and RL) were stored at the fourth instar larval stage.

DNA extraction and library preparation

Whole genomic DNA was extracted from each individual mosquito using Qiagen DNA Blood and Tissue kit (Venlo, Limburg, NL) for 2010 and 2013 samples, using Roche High Pure^{TM} PCR Template Preparation Kits (Roche Molecular Systems, Inc., Pleasanton, CA, USA) for 2018 samples. *Wolbachia* infection status was confirmed by a diagnostic qPCR test as outlined elsewhere [62]. The concentration of extracted individual DNA was measured using Quantitation Qubit™ 1X dsDNA HS Assay Kit (Invitrogen Life Technologies USA). Samples from each of the six populations (Table 4) were pooled prior to sequencing based on an equal amount of DNA from each individual. Each population was sent for whole genome sequencing with > 50 depth via Illumina Hiseq2500 using 100 bp paired read chemistry for GV10 and GV13, and 150 bp paired read chemistry for GV18, EH, YK and RL libraries.

Raw sequences were trimmed using Trimmomatic v. 0.39 to truncate and remove low quality reads; the reads with a phred score above 20 and length above 70 were kept. The reference genome AaegL5.0 [30] was indexed and reads were aligned to the reference by bowtie2 v. 2.3.4.3 with the very-sensitive-local mode [63]. Samtools v.1.9 was used with default parameters to sort, mark and generate pileup files requiring a minimal mapping quality of 20.

Estimation of genome variation

We investigated patterns of genetic variation within populations using PoPoolation v. 1.2.2 [32] with the genomic annotation file from the reference AaegL5.0. We calculated Tajima's pi (nucleotide diversity π) for each population at 10 kbp non-overlapping windows when the coverage was not less than 20. Windows with low coverage generated no values and were neglected before adjusting the shape of lines across each chromosome by a LOESS smooth curve [64]. We also calculated Tajima's D for each population at 10 kbp non-overlapping windows and at the gene level with a minimal coverage of 20. The value of Tajima's D was calculated from allele frequencies in selected regions and was used to detect selection direction. Under a standard neutral model with no change in population size, a strongly negative Tajima's D value can indicate directional selection removing variation, while a strongly positive value can indicate balancing selection maintaining variation, with 0 reflecting an absence of selection.

For the genetic variation between populations, we used PoPoolation2 v. 1.201 [65] to obtain allele frequency differences for each SNP. The SNPs were then filtered by the following parameters before further analyses: locus coverage > 50 in all populations and an average minor allele frequency (MAF) more than 0.1 [66]. We also obtained pairwise Fst values for non-overlapping 100 kbp windows and for gene sets. We then undertook a principal component analysis (PCA) generated by the prcomp function and package ggbiplot in R [67] based on: 1) MAF across the SNPs after initial filtering as mentioned above; 2) pairwise Fst values from 100 kbp non-overlapping windows to indicate genetic distance patterns across the genome (genome level) [68]; and 3) pairwise Fst for genes (gene level). These Fst estimates were then used to assess patterns of similarity among samples with the same *Wolbachia* infection status in a pairwise comparison, and the same geographic distance in a pairwise comparison.

We further investigated isolation by distance patterns among 2018 samples from the four locations by computing $Fst^*=Fst/(1 - Fst)$ [69] based on the averaged pairwise Fst values from 100k bp non-overlapping windows. A geographic distance matrix was built based on the natural log transformation of the shortest road distance between the sampled locations as mosquito movement would be mostly by road transport [70]. We then looked for Fst patterns that might be related to this distance measure and ran a Mantel test through the ade4 package in R to test the relationship between genetic distance and geographic distance [71], with only 2018 populations included.

Identification of outliers potentially associated with *Wolbachia*

We used two models from a Bayesian outlier approach BayPass v. 2.2 [35] to identify outliers associated with *Wolbachia* infection. Firstly, we used standard covariate model [35, 72], which requires a file providing values of each covariate to produce the Bayes Factor (BF), the ratio of the likelihood of posterior and prior hypotheses, which can quantify the probability of a candidate SNP being under selection [73]. Bayes Factors were exported in dB units ($BF^* = 10 \times \log10(BF)$); the association with environmental variance was considered "substantial" when $BF^* > 5$, "strongly-

selected" when BF* > 10 and "decisively-selected" when BF^* > 13 following Jeffreys [73]. We modelled *Wolbachia* infection status as a binary covariate by setting each infected sample as 1 and each uninfected sample as 0. The measures of BFs are based on an Importance Sampling Approximation, which is unstable for single runs in particular when the number of populations is small, so we averaged the BF values of three runs with different seeds for the random number generators following the suggestion in the manual of BayPass [35]. SNPs were considered as "substantial" outliers if the average BF* >= 5 and "strongly-selected" outliers if the average BF* >= 10. The introduction of linear relationships in this model, however, can have high false positives from sampling noise, as the relationships of allele frequency among multiple populations are influenced by genetic drift and migration, and are not statistically independent with each other [36]. Based on this model, we made Manhattan plots for SNPs with average BF* > 0 to show the potential effects of the *Wolbachia* infection at the genome level; SNPs that could not be assigned a position on one of the three autosomes were discarded.

Secondly, we used a BayPass core model to identify outliers from comparisons between GV10 and GV13, GV10 and GV18, given that Gordonvale was the only location where we had samples at time points before and after the release. The infection of *Wolbachia* was considered the main variable across time. An X^TX algorithm approach was used in this model, which was analogous to an Fst comparison [35, 36]. The X^TX value was used to identify selection pressure, with higher values under positive selection and smaller values under balancing selection [36]. In this model, we considered SNPs with X^TX values greater than the 90 % threshold or 95% threshold in both the GV10 and GV13 comparison and the GV10 and GV18 comparison as candidates for intersecting with the "substantial" SNPs and "strongly-selected" SNPs from the first analysis. This model on the other hand, is unable to exclude the noise from gene flow or genetic drift due to the lack of duplicates, which can also cause false positives. We therefore considered the intersection of outliers from the above two models as SNPs potentially associated with *Wolbachia* infection. These SNPs were then matched with GTF annotation file from NCBI to obtain a list of outlier genes.

Pathway analysis and gene ontology enrichment analysis

We considered SNPs from the intersection of the two BayPass models above as outliers potentially under selection and searched the open reading frame (ORF) to obtain a list of potentially important genes within a 100 kbp region around each SNP (50 kbp upstream and 50 kbp downstream) [74]. These genes were then searched through the KEGG pathway database for interpretation. We then used BLAST search (version 2.9.0 +) in the UniProtKB/Swiss-Prot database [75] through system Entrez to identify homologous proteins in *Drosophila melanogaster*. Only best matches were retained and further filtered with an e-value cut off 1.0E-10 and > 60% identity [76, 77]. GO enrichment analysis was then undertaken in the R packages clusterProfiler

[78] and DOSE [79], with a significant false positive rate and false discovery rate cut-off of 0.05.

Additional files

Additional file 1. Genetic and geographic distance matrices. The genetic distance matrix from genome pairwise 100 kbp Fst value, transformed by Fst*=Fst/(1 − Fst), is above the diagonal, the nature log transformed geographic distance (km) is below the diagonal.

	GV10	GV13	GV18	EH	RL	YK
GV10	0	0.045	0.040	0.056	0.059	0.064
GV13	0	0	0.057	0.063	0.067	0.068
GV18	0	0	0	0.052	0.057	0.059
EH	3.243	3.243	3.243	0	0.051	0.054
RL	3.547	3.547	3.547	2.526	0	0.057
YK	3.694	3.694	3.694	2.821	2.773	0

Additional file 2. IDs of 187 genes where the 950 "substantial" outliers that potentially impacted by *Wolbachia w*Mel infection distributed within.

Additional file 3. IDs of 1436 genes within the 100 kbp regions around the 950 outliers that potentially impacted by *Wolbachia w*Mel infection.

Additional file 4. Pathways in KEGG database identified from genes potentially impacted by *Wolbachia w*Mel infection

Additional file 5. Homologues proteins of *Drosophila melanogaster* in GO enrichment analysis.

Additional file 6. Enriched biological process terms.

Additional file 7. Enriched cellular component terms.

Additional file 8. Enriched molecular function terms.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The raw data used during the current study belongs to corresponding authors. All data generated during this study are included within this article and its additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the National Health and Medical Research Council (1132412, 1118640, www.nhmrc.gov.au). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

AH and TS conceived the study. ML implemented the study, performed data analysis and wrote the first draft of the manuscript. AH, TS, PAR, QY and LS revised the manuscript. JC and QY provided guidance in data analysis. ML, TS, LS, QY and PAR contributed to sample preparation and/or sequencing. AH secured financial support. All authors read and approved the final manuscript.

Acknowledgements

We thank Gordana Rasic for her early contribution to the methodology for initial sequencing of the Gordonvale populations. We also thank Moshe Jasper for his advice on data visualization. This research was supported by use of the Nectar Research Cloud, a collaborative Australian research platform supported by the NCRIS-funded Australian Research Data Commons (ARDC).

References

1. Nazni WA, Hoffmann AA, NoorAfizah A, Cheong YL, Mancini MV, Golding N, Kamarul GM, Arif MA, Thohir H, NurSyamimi H: **Establishment of**

***Wolbachia* strain *w*AlbB in Malaysian populations of *Aedes aegypti* for dengue control**. *Curr Biol* 2019, **29**(24):4241-4248.

2. Garcia GdA, Sylvestre G, Aguiar R, da Costa GB, Martins AJ, Lima JBP, Petersen MT, Lourenço-de-Oliveira R, Shadbolt MF, Rašić G *et al*: **Matching the genetics of released and local *Aedes aegypti* populations is critical to assure *Wolbachia* invasion**. *PLoS neglected tropical diseases* 2019, **13**(1):e0007023.
3. Ryan PA, Turley AP, Wilson G, Hurst TP, Retzki K, Brown-Kenyon J, Hodgson L, Kenny N, Cook H, Montgomery BL: **Establishment of *w*Mel *Wolbachia* in *Aedes aegypti* mosquitoes and reduction of local dengue transmission in Cairns and surrounding locations in northern Queensland, Australia**. *Gates Open Res* 2019, **3**.
4. Van den Hurk AF, Hall-Mendelin S, Pyke AT, Frentiu FD, McElroy K, Day A, Higgs S, O'Neill SL: **Impact of *Wolbachia* on infection with chikungunya and yellow fever viruses in the mosquito vector *Aedes aegypti***. *PLoS neglected tropical diseases* 2012, **6**(11):e1892.
5. Aliota MT, Peinado SA, Velez ID, Osorio JE: **The *w*Mel strain of *Wolbachia* reduces transmission of Zika virus by *Aedes aegypti***. *Sci Rep* 2016, **6**:28792.
6. Walker T, Johnson PH, Moreira LA, Iturbe-Ormaetxe I, Frentiu FD, McMeniman CJ, Leong YS, Dong Y, Axford J, Kriesner P *et al*: **The *w*Mel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations**. *Nature* 2011, **476**(7361):450-453.
7. Hoffmann AA, Montgomery B, Popovici J, Iturbe-Ormaetxe I, Johnson P, Muzzi F, Greenfield M, Durkan M, Leong Y, Dong Y: **Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission**. *Nature* 2011, **476**(7361):454-457.
8. Turelli M, Hoffmann AA: **Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations**. *Genetics* 1995, **140**(4):1319-1338.
9. Telschow A, Hammerstein P, Werren JH: **Effects of *Wolbachia* on genetic divergence between populations: mainland-island model**. *Integr Comp Biol* 2002, **42**(2):340-351.
10. Telschow A, Hammerstein P, Werren JH: **The effect of *Wolbachia* on genetic divergence between populations: models with two-way migration**. *Am Nat* 2002, **160 Suppl 4**:S54-66.
11. Kraaijeveld K, Franco P, de Knijff P, Stouthamer R, van Alphen JJ: **Clonal genetic variation in a *Wolbachia*-infected asexual wasp: horizontal transmission or historical sex?** *Molecular ecology* 2011, **20**(17):3644-3652.
12. Hale LR, Hoffmann AA: **Mitochondrial DNA polymorphism and cytoplasmic incompatibility in natural populations of *Drosophila simulans***. *Evolution* 1990, **44**(5):1383-1386.
13. Yeap HL, Rašić G, Endersby-Harshman NM, Lee SF, Arguni E, Le Nguyen H, Hoffmann AA: **Mitochondrial DNA variants help monitor the dynamics**

of *Wolbachia* invasion into host populations. *Heredity (Edinb)* 2016, **116**(3):265-276.
14. Hoffmann AA, Iturbe-Ormaetxe I, Callahan AG, Phillips BL, Billington K, Axford JK, Montgomery B, Turley AP, O'Neill SL: **Stability of the *w*Mel *Wolbachia* infection following invasion into *Aedes aegypti* populations**. *PLoS neglected tropical diseases* 2014, **8**(9):e3115.
15. McMeniman CJ, Lane RV, Cass BN, Fong AWC, Sidhu M, Wang Y-F, O'Neill SL: **Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti***. *Science* 2009, **323**(5910):141-144.
16. Ross PA, Wiwatanaratanabutr I, Axford JK, White VL, Endersby-Harshman NM, Hoffmann AA: ***Wolbachia* infections in *Aedes aegypti* differ markedly in their response to cyclical heat stress**. *PLoS Pathog* 2017, **13**(1):e1006006.
17. Bull JJ, Turelli M: ***Wolbachia* versus dengue: Evolutionary forecasts**. *Evol Med Public Health* 2013, **2013**(1):197-207.
18. Ross P, Hoffmann A: **Continued susceptibility of the *w*Mel *Wolbachia* infection in *Aedes aegypti* to heat stress following field deployment and selection**. *Insects* 2018, **9**(3):78.
19. Ford SA, Allen SL, Ohm JR, Sigle LT, Sebastian A, Albert I, Chenoweth SF, McGraw EA: **Selection on *Aedes aegypti* alters *Wolbachia*-mediated dengue virus blocking and fitness**. *Nat Microbiol* 2019, **4**(11):1832-1839.
20. Ritchie SA, Townsend M, Paton CJ, Callahan AG, Hoffmann AA: **Application of *w*MelPop *Wolbachia* strain to crash local populations of *Aedes aegypti***. *PLoS neglected tropical diseases* 2015, **9**(7):e0003930.
21. Weeks AR, Turelli M, Harcombe WR, Reynolds KT, Hoffmann AA: **From parasite to mutualist: rapid evolution of *Wolbachia* in natural populations of *Drosophila***. *PLoS biology* 2007, **5**(5):e114.
22. Hornett EA, Charlat S, Duplouy AM, Davies N, Roderick GK, Wedell N, Hurst GD: **Evolution of male-killer suppression in a natural population**. *PLoS Biol* 2006, **4**(9):e283.
23. Carrington LB, Hoffmann AA, Weeks AR: **Monitoring long-term evolutionary changes following *Wolbachia* introduction into a novel host: the *Wolbachia* popcorn infection in *Drosophila simulans***. *Proc Biol Sci* 2010, **277**(1690):2059-2068.
24. Ross PA, Turelli M, Hoffmann AA: **Evolutionary Ecology of *Wolbachia* Releases for Disease Control**. *Annu Rev Genet* 2019, **53**:93-116.
25. Huang B, Yang Q, Hoffmann AA, Ritchie SA, van den Hurk AF, Warrilow D: ***Wolbachia* genome stability and mtDNA variants in *Aedes aegypti* field populations eight years after release**. *iScience* 2020:101572.
26. Frentiu FD, Zakir T, Walker T, Popovici J, Pyke AT, van den Hurk A, McGraw EA, O'Neill SL: **Limited dengue virus replication in field-collected *Aedes aegypti* mosquitoes infected with *Wolbachia***. *PLoS neglected tropical diseases* 2014, **8**(2).
27. Carrington LB, Tran BCN, Le NTH, Luong TTH, Nguyen TT, Nguyen PT, Nguyen CVV, Nguyen HTC, Vu TT, Vo LT *et al*: **Field- and clinically**

derived estimates of *Wolbachia*-mediated blocking of dengue virus transmission potential in *Aedes aegypti* mosquitoes. *Proc Natl Acad Sci U S A* 2018, **115**(2):361-366.
28. Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, Loftus B, Xi Z, Megy K, Grabherr M: **Genome sequence of *Aedes aegypti*, a major arbovirus vector**. *Science* 2007, **316**(5832):1718-1723.
29. Maringer K, Yousuf A, Heesom KJ, Fan J, Lee D, Fernandez-Sesma A, Bessant C, Matthews DA, Davidson AD: **Proteomics informed by transcriptomics for characterising active transposable elements and genome annotation in *Aedes aegypti***. *BMC Genomics* 2017, **18**(1):101-101.
30. Matthews BJ, Dudchenko O, Kingan SB, Koren S, Antoshechkin I, Crawford JE, Glassford WJ, Herre M, Redmond SN, Rose NH *et al*: **Improved reference genome of *Aedes aegypti* informs arbovirus vector control**. *Nature* 2018, **563**(7732):501-507.
31. Feschotte C: **Transposable elements and the evolution of regulatory networks**. *Nat Rev Genet* 2008, **9**(5):397-405.
32. Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C: **PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals**. *PLoS One* 2011, **6**(1):e15925-e15925.
33. Lee Y, Schmidt H, Collier TC, Conner WR, Hanemaaijer MJ, Slatkin M, Marshall JM, Chiu JC, Smartt CT, Lanzaro GC *et al*: **Genome-wide divergence among invasive populations of *Aedes aegypti* in California**. *BMC Genomics* 2019, **20**(1):204.
34. Timoshevskiy VA, Kinney NA, Becky Sd, Mao C, Tu Z, Severson DW, Sharakhov IV, Sharakhova MV: **Genomic composition and evolution of *Aedes aegypti* chromosomes revealed by the analysis of physically mapped supercontigs**. *BMC Biol* 2014, **12**(1):27.
35. Gautier M: **Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates**. *Genetics* 2015, **201**(4):1555-1579.
36. Günther T, Coop G: **Robust identification of local adaptation from allele frequencies**. *Genetics* 2013, **195**(1):205-220.
37. Behura SK, Gomez-Machorro C, Harker BW, deBruyn B, Lovin DD, Hemme RR, Mori A, Romero-Severson J, Severson DW: **Global cross-talk of genes of the mosquito *Aedes aegypti* in response to dengue virus infection**. *PLoS neglected tropical diseases* 2011, **5**(11):e1385.
38. Bariami V, Jones CM, Poupardin R, Vontas J, Ranson H: **Gene amplification, ABC transporters and cytochrome P450s: unraveling the molecular basis of pyrethroid resistance in the dengue vector, *Aedes aegypti***. *PLoS neglected tropical diseases* 2012, **6**(6):e1692.
39. Stevenson BJ, Pignatelli P, Nikou D, Paine MJ: **Pinpointing P450s associated with pyrethroid metabolism in the dengue vector, *Aedes aegypti*: developing new tools to combat insecticide resistance**. *PLoS neglected tropical diseases* 2012, **6**(3):e1595.

40. Corena M, Seron TJ, Lehman HK, Ochrietor JD, Kohn A, Tu C, Linser PJ: **Carbonic anhydrase in the midgut of larval *Aedes aegypti*: cloning, localization and inhibition**. *J Exp Biol* 2002, **205**(Pt 5):591-602.
41. del Pilar Corena M, VanEkeris L, Salazar MI, Bowers D, Fiedler MM, Silverman D, Tu C, Linser PJ: **Carbonic anhydrase in the adult mosquito midgut**. *J Exp Biol* 2005, **208**(Pt 17):3263-3273.
42. Pierceall WE, Li C, Biran A, Miura K, Raikhel AS, Segraves WA: **E75 expression in mosquito ovary and fat body suggests reiterative use of ecdysone-regulated hierarchies in development and reproduction**. *Mol Cell Endocrinol* 1999, **150**(1-2):73-89.
43. van Heusden MC, Thompson F, Dennis J: **Biosynthesis of *Aedes aegypti* lipophorin and gene expression of its apolipoproteins**. *Insect Biochem Mol Biol* 1998, **28**(10):733-738.
44. Bartholomay LC, Cho WL, Rocheleau TA, Boyle JP, Beck ET, Fuchs JF, Liss P, Rusch M, Butler KM, Wu RC *et al*: **Description of the transcriptomes of immune response-activated hemocytes from the mosquito vectors *Aedes aegypti* and Armigeres subalbatus**. *Infect Immun* 2004, **72**(7):4114-4126.
45. Endersby NM, Hoffmann AA, White VL, Lowenstein S, Ritchie S, Johnson PH, Rapley LP, Ryan PA, Nam VS, Yen NT *et al*: **Genetic structure of *Aedes aegypti* in Australia and Vietnam revealed by microsatellite and exon primed intron crossing markers suggests feasibility of local control options**. *J Med Entomol* 2009, **46**(5):1074-1083.
46. Turelli M, Hoffmann AA, McKechnie SW: **Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations**. *Genetics* 1992, **132**(3):713-723.
47. Waterhouse RM, Kriventseva EV, Meister S, Xi Z, Alvarez KS, Bartholomay LC, Barillas-Mury C, Bian G, Blandin S, Christensen BM *et al*: **Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes**. *Science* 2007, **316**(5832):1738-1743.
48. Pan X, Pike A, Joshi D, Bian G, McFadden MJ, Lu P, Liang X, Zhang F, Raikhel AS, Xi Z: **The bacterium *Wolbachia* exploits host innate immunity to establish a symbiotic relationship with the dengue vector mosquito *Aedes aegypti***. *The ISME journal* 2018, **12**(1):277-288.
49. Xi Z, Ramirez JL, Dimopoulos G: **The *Aedes aegypti* toll pathway controls dengue virus infection**. *PLoS Pathog* 2008, **4**(7):e1000098-e1000098.
50. Kambris Z, Blagborough AM, Pinto SB, Blagrove MSC, Godfray HCJ, Sinden RE, Sinkins SP: ***Wolbachia* stimulates immune gene expression and inhibits plasmodium development in *Anopheles gambiae***. *PLoS Pathog* 2010, **6**(10):e1001143-e1001143.
51. Kambris Z, Cook PE, Phuc HK, Sinkins SP: **Immune Activation by Life-Shortening Wolbachia and Reduced Filarial Competence in Mosquitoes**. *Science* 2009, **326**(5949):134-136.

52. Rancès E, Ye YH, Woolfit M, McGraw EA, O'Neill SL: **The relative importance of innate immune priming in *Wolbachia*-mediated dengue interference**. *PLoS Pathog* 2012, **8**(2):e1002548.
53. Coelho A, Fraichard S, Le Goff G, Faure P, Artur Y, Ferveur JF, Heydel JM: **Cytochrome P450-dependent metabolism of caffeine in *Drosophila melanogaster***. *PLoS One* 2015, **10**(2):e0117328.
54. Chung H, Sztal T, Pasricha S, Sridhar M, Batterham P, Daborn PJ: **Characterization of *Drosophila melanogaster* cytochrome P450 genes**. *Proc Natl Acad Sci U S A* 2009, **106**(14):5731-5736.
55. McGraw E, Merritt D, Droller J, O'Neill SL: ***Wolbachia* density and virulence attenuation after transfer into a novel host**. *Proc Natl Acad Sci U S A* 2002, **99**(5):2918-2923.
56. Turelli M: **Evolution of incompatibility-inducing microbes and their hosts**. *Evolution* 1994, **48**(5):1500-1513.
57. Terradas G, Allen SL, Chenoweth SF, McGraw EA: **Family level variation in *Wolbachia*-mediated dengue virus blocking in *Aedes aegypti***. *Parasit Vectors* 2017, **10**(1):622.
58. Echaubard P, Duron O, Agnew P, Sidobre C, Noël V, Weill M, Michalakis Y: **Rapid evolution of *Wolbachia* density in insecticide resistant *Culex pipiens***. *Heredity (Edinb)* 2010, **104**(1):15-19.
59. Carrington LB, Leslie J, Weeks AR, Hoffmann AA: **The popcorn *Wolbachia* infection of *Drosophila melanogaster*: can selection alter i longevity effects?** *Evolution* 2009, **63**(10):2648-2657.
60. Schmidt TL, Barton NH, Rašić G, Turley AP, Montgomery BL, Iturbe-Ormaetxe I, Cook PE, Ryan PA, Ritchie SA, Hoffmann AA: **Local introduction and heterogeneous spatial spread of dengue-suppressing *Wolbachia* through an urban population of *Aedes aegypti***. *PLoS Biol* 2017, **15**(5):e2001894.
61. Jasper M, Schmidt TL, Ahmad NW, Sinkins SP, Hoffmann AA: **A genomic approach to inferring kinship reveals limited intergenerational dispersal in the yellow fever mosquito**. *Mol Ecol Resour* 2019.
62. Lee SF, White VL, Weeks AR, Hoffmann AA, Endersby NM: **High-throughput PCR assays to monitor *Wolbachia* infection in the dengue mosquito (*Aedes aegypti*) and *Drosophila simulans***. *Appl Environ Microbiol* 2012, **78**(13):4740-4743.
63. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2**. *Nat Methods* 2012, **9**(4):357-359.
64. Wickham H: **ggplot2: elegant graphics for data analysis**: Springer; 2016.
65. Kofler R, Pandey RV, Schlötterer C: **PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq)**. *Bioinformatics* 2011, **27**(24):3435-3436.
66. Asif H, Alliey-Rodriguez N, Keedy S, Tamminga CA, Sweeney JA, Pearlson G, Clementz BA, Keshavan MS, Buckley P, Liu C *et al*: **GWAS significance**

thresholds for deep phenotyping studies can depend upon minor allele frequencies and sample size. *Mol Psychiatry* 2020.
67. Team RC: **R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria**. In.; 2019.
68. Hivert V, Leblois R, Petit EJ, Gautier M, Vitalis R: **Measuring Genetic Differentiation from Pool-seq Data**. *Genetics* 2018, **210**(1):315-330.
69. Rousset F: **Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance**. *Genetics* 1997, **145**(4):1219-1228.
70. Medley KA, Jenkins DG, Hoffman EA: **Human-aided and natural dispersal drive gene flow across the range of an invasive mosquito**. *Molecular ecology* 2015, **24**(2):284-295.
71. Diniz-Filho JAF, Soares TN, Lima JS, Dobrovolski R, Landeiro VL, de Campos Telles MP, Rangel TF, Bini LM: **Mantel test in population genetics**. *Genet Mol Biol* 2013, **36**(4):475-485.
72. Coop G, Witonsky D, Di Rienzo A, Pritchard JK: **Using environmental correlations to identify loci underlying local adaptation**. *Genetics* 2010, **185**(4):1411-1423.
73. Jeffreys H: **The theory of probability**. Oxford.: Oxford University Press; 1961.
74. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features**. *Bioinformatics* 2010, **26**(6):841-842.
75. Camon E, Magrane M, Barrell D, Lee V, Dimmer E, Maslen J, Binns D, Harte N, Lopez R, Apweiler R: **The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology**. *Nucleic Acids Res* 2004, **32**(suppl_1):D262-D266.
76. Pearson WR: **An introduction to sequence similarity ("homology") searching**. *Curr Protoc Bioinformatics* 2013, **Chapter 3**:Unit3.1.
77. Tian W, Skolnick J: **How well is enzyme function conserved as a function of pairwise sequence identity?** *J Mol Biol* 2003, **333**(4):863-882.
78. Yu G, Wang L-G, Han Y, He Q-Y: **clusterProfiler: an R package for comparing biological themes among gene clusters**. *OMICS* 2012, **16**(5):284-287.
79. Yu G, Wang L-G, Yan G-R, He Q-Y: **DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis**. *Bioinformatics* 2015, **31**(4):608-609.