



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Lau, MJ;Schmidt, TL;Yang, Q;Chung, J;Sankey, L;Ross, PA;Hoffmann, AA

Title:

Genetic stability of *Aedes aegypti* populations following invasion by wMel Wolbachia

Date:

2021-12-01

Citation:

Lau, M. J., Schmidt, T. L., Yang, Q., Chung, J., Sankey, L., Ross, P. A. & Hoffmann, A. A. (2021). Genetic stability of *Aedes aegypti* populations following invasion by wMel Wolbachia. *BMC Genomics*, 22 (1), <https://doi.org/10.1186/s12864-021-08200-1>.

Persistent Link:

<https://hdl.handle.net/11343/295941>

License:

CC BY

RESEARCH

Open Access



# Genetic stability of *Aedes aegypti* populations following invasion by *wMel Wolbachia*

Meng-Jia Lau<sup>1\*</sup>, Thomas L. Schmidt<sup>1\*</sup>, Qiong Yang<sup>1</sup>, Jessica Chung<sup>1,2</sup>, Lucien Sankey<sup>1</sup>, Perran A. Ross<sup>1</sup> and Ary A. Hoffmann<sup>1\*</sup>

## Abstract

**Background:** *Wolbachia wMel* is the most commonly used strain in rear and release strategies for *Aedes aegypti* mosquitoes that aim to inhibit the transmission of arboviruses such as dengue, Zika, Chikungunya and yellow fever. However, the long-term establishment of *wMel* in natural *Ae. aegypti* populations raises concerns that interactions between *Wolbachia wMel* 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 wMel* 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:** *Ae. aegypti* populations remained geographically distinct after *Wolbachia wMel* 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, Genome-wide association study

## Background

*Wolbachia* are bacteria that live inside the cells of many insects and induce 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 wMel*, which was transferred artificially from *Drosophila melanogaster* into *Ae. aegypti* [6], was first

\*Correspondence: mengjial2@student.unimelb.edu.au; tom.schmidt@unimelb.edu.au; ary@unimelb.edu.au

<sup>1</sup> Pest and Environmental Adaptation Research Group, Bio21 Institute and the School of BioSciences, The University of Melbourne, Parkville, Victoria, Australia

Full list of author information is available at the end of the article



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* *wMel*-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 about one 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 transacted into new hosts [24].

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, in *Ae. aegypti*, there are challenges in characterizing genome changes after *Wolbachia* *wMel* 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 [25–27]. TEs might also enhance rates of evolution, given that they are involved with gene regulation, and increase genome plasticity [28]. Moreover, other environmental factors in field-collected samples may be important, and the

impacts of gene flow following the activity of *Wolbachia* release will increase the difficulty of outlier analysis. 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>).

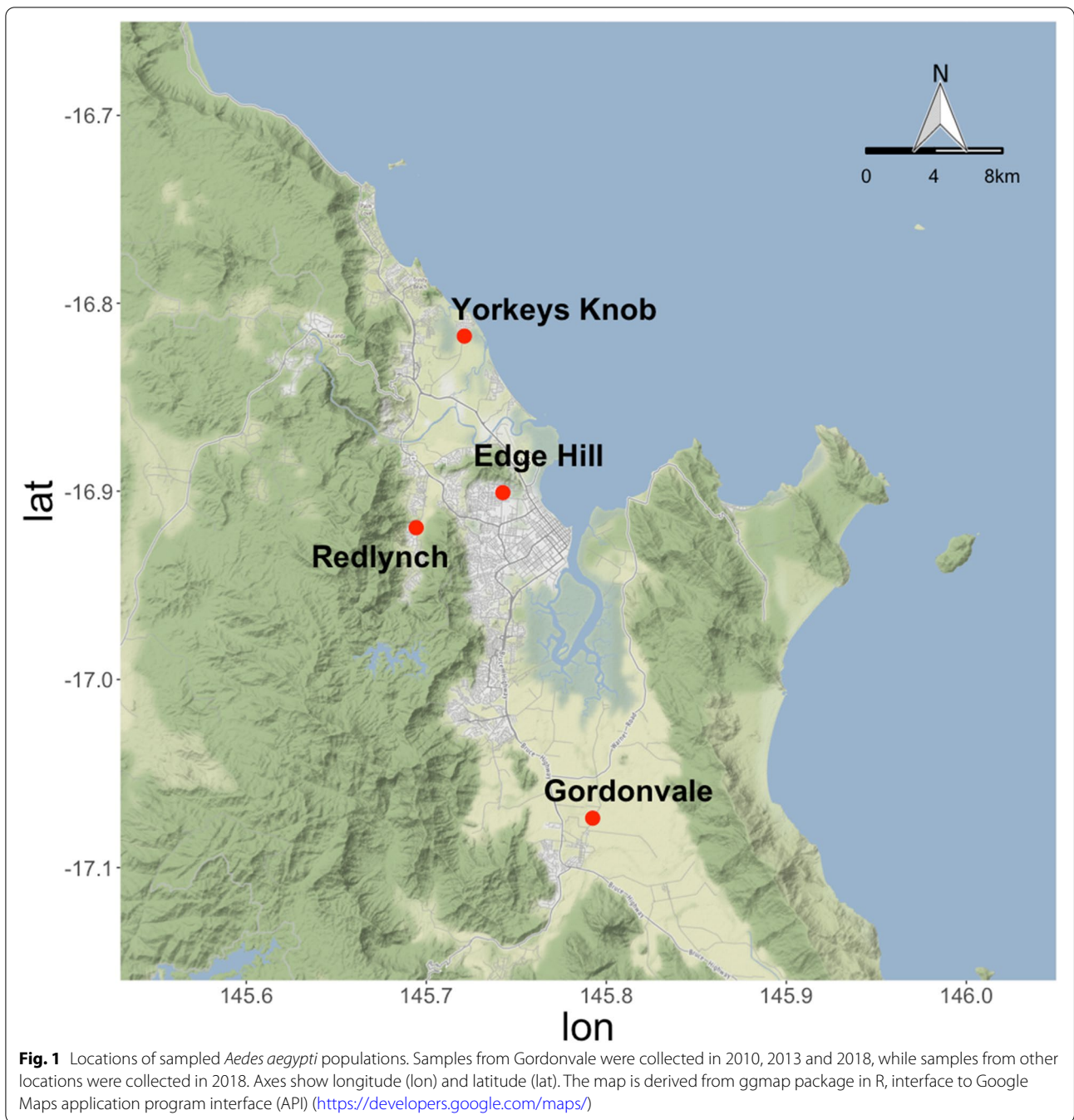
In this study, we analyze pooled whole genome sequencing (WGS) data of mosquitoes from Gordonvale, Australia, where releases first took place, covering three different time points from the pre- and post-release phase. As a comparison, we also sequenced samples from the nearby Cairns area in Edge Hill,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

## Results

### Genetic variation in *Aedes aegypti* populations

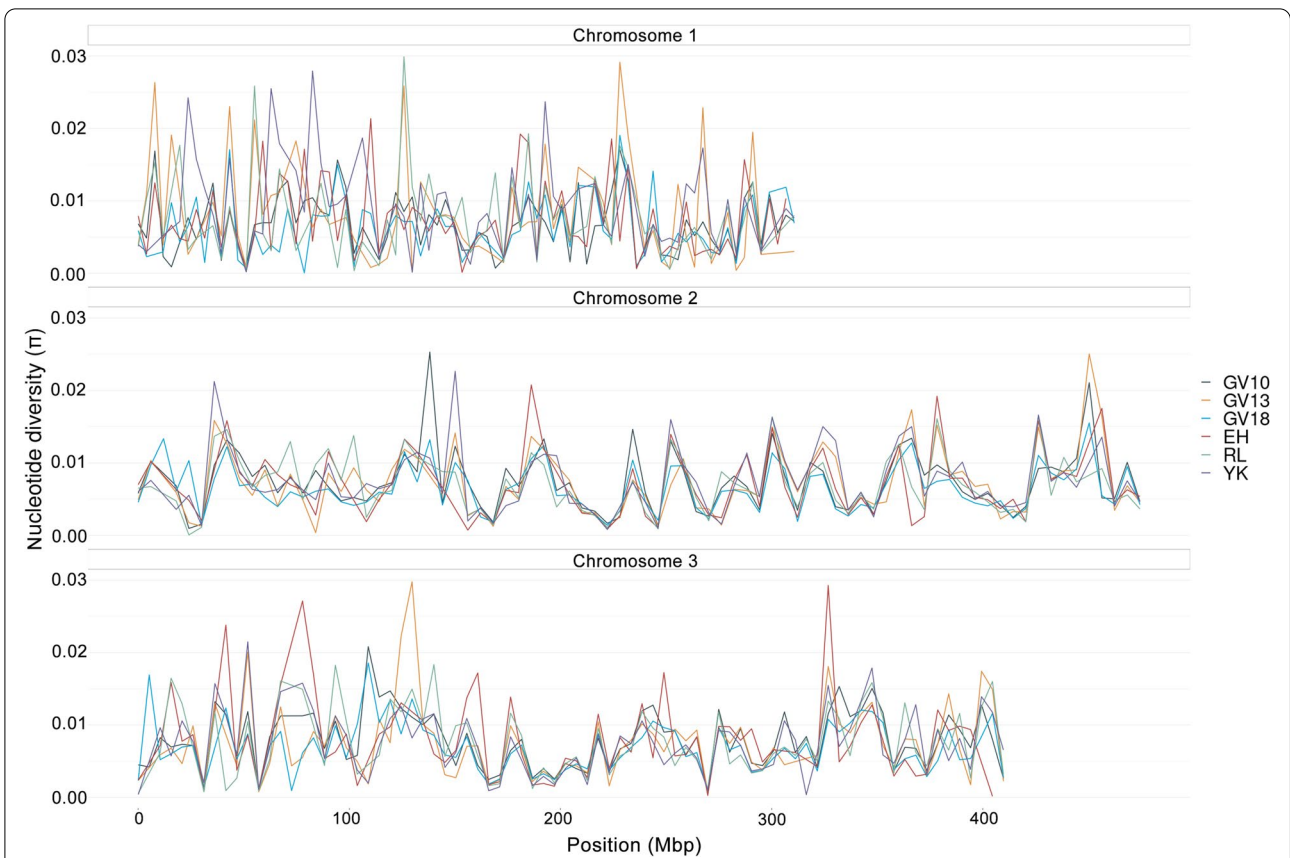
*Aedes aegypti* were collected from four sites around Cairns, Australia, at different times pre- and post-*Wolbachia* releases (Fig. 1, Table 1). We investigated patterns of genetic variation within populations and obtained Tajima's  $\pi$  (nucleotide diversity  $\pi$ ) for each population at 10 kbp non-overlapping windows. Differences in  $\pi$  among populations depended on chromosomal location. Near the center of each chromosome, diversities were low and similar for all populations (Fig. 2). In areas far from the centromere, variation between populations was large. This pattern has been observed in previous sequencing studies [27, 29], which may result from a high proportion of TEs and satellites in non-coding regions. 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 [30].

We calculated Tajima's D for each population at 10 kbp non-overlapping windows and at the gene level. We found that the density distributions of values were

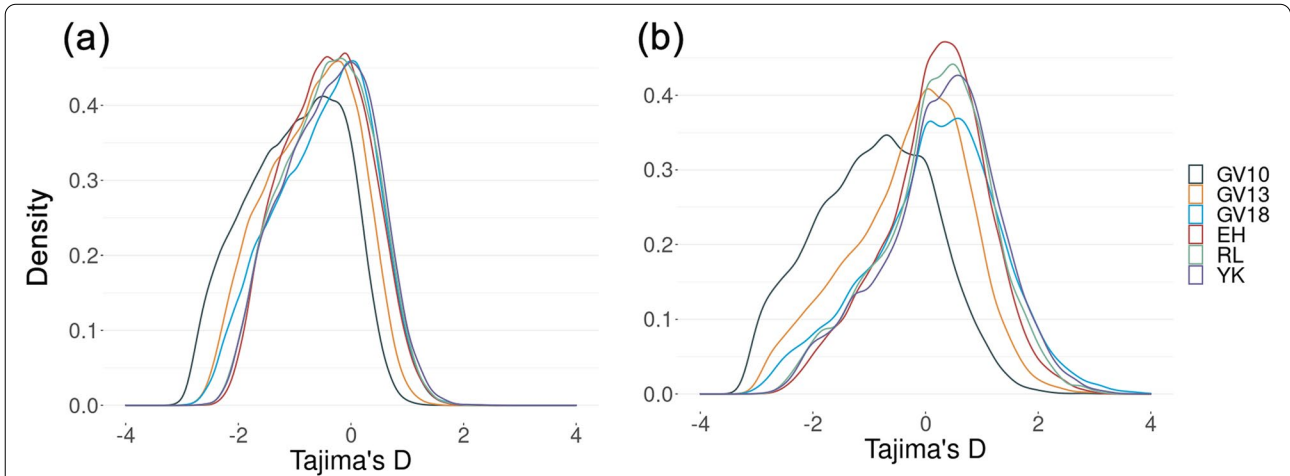


**Table 1** Summary of *Aedes aegypti* collections and designations of samples used in comparisons

Population name	Collecting year	Location	<i>Wolbachia</i> 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 ( $\pi$ ). 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



**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

similar between populations, except for the Gordonvale pre-release (GV10) sample 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), Additional file 1). The four 2018 populations converge regardless of their *Wolbachia* infection 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.

#### Geographic segregation of *Aedes aegypti* populations

We investigated genetic distances from the average of pairwise  $F_{st}$  (Fixation index) values through 100 kbp non-overlapping windows. The temporally-separated Gordonvale samples tended to have lower pairwise  $F_{st}$  than population pairs from different locations (Table 2). For isolation by distance (IBD) test, we also obtained a geographic distance matrix with the four 2018 populations, which was built based on the natural log transformation of the shortest road distance between the sampled locations (Table 2). Genetic distance had a weak but non-significant correlation with natural log transformed geographic distance in a Mantel test ( $r = 0.66, p = 0.12$ ).

There were 461,067 single-nucleotide polymorphisms (SNPs) left after filtering with minimal depth of 50 in all populations and an average minor allele frequency (MAF) > 0.1. A principal components analysis (PCA) based on these SNPs showed the first two PCs accounted for 24.2 and 22.9% of the variance respectively (Fig. 4a). The three temporally-separated Gordonvale samples fell out together, but GV10 was closer to GV18 than to GV13. We also ran PCAs on pairwise  $F_{st}$  differences in 100 kbp non-overlapping windows across the genome (Fig. 4b, d) and 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 clustered geographically (Fig. 4b, c). Comparisons between the Gordonvale samples and Yorkkeys Knob, Edge Hill, and Redlynch showed consistent

clustering by location, and stronger clustering at the gene level (Fig. 4c) than across the genome (Fig. 4b).

#### Bayesian models to identify outliers potentially associated with *Wolbachia*

To investigate potential selection associated with *Wolbachia*, we used 461,067 SNPs from the above filtering process and used two Bayesian models from BayPass v. 2.2 [31] 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  $BF^*$  (Bayes Factor (BF) in dB units) > 5, and 391 showing “strongly-selected” signature of selection with an average  $BF^* > 10$  (Fig. 5).

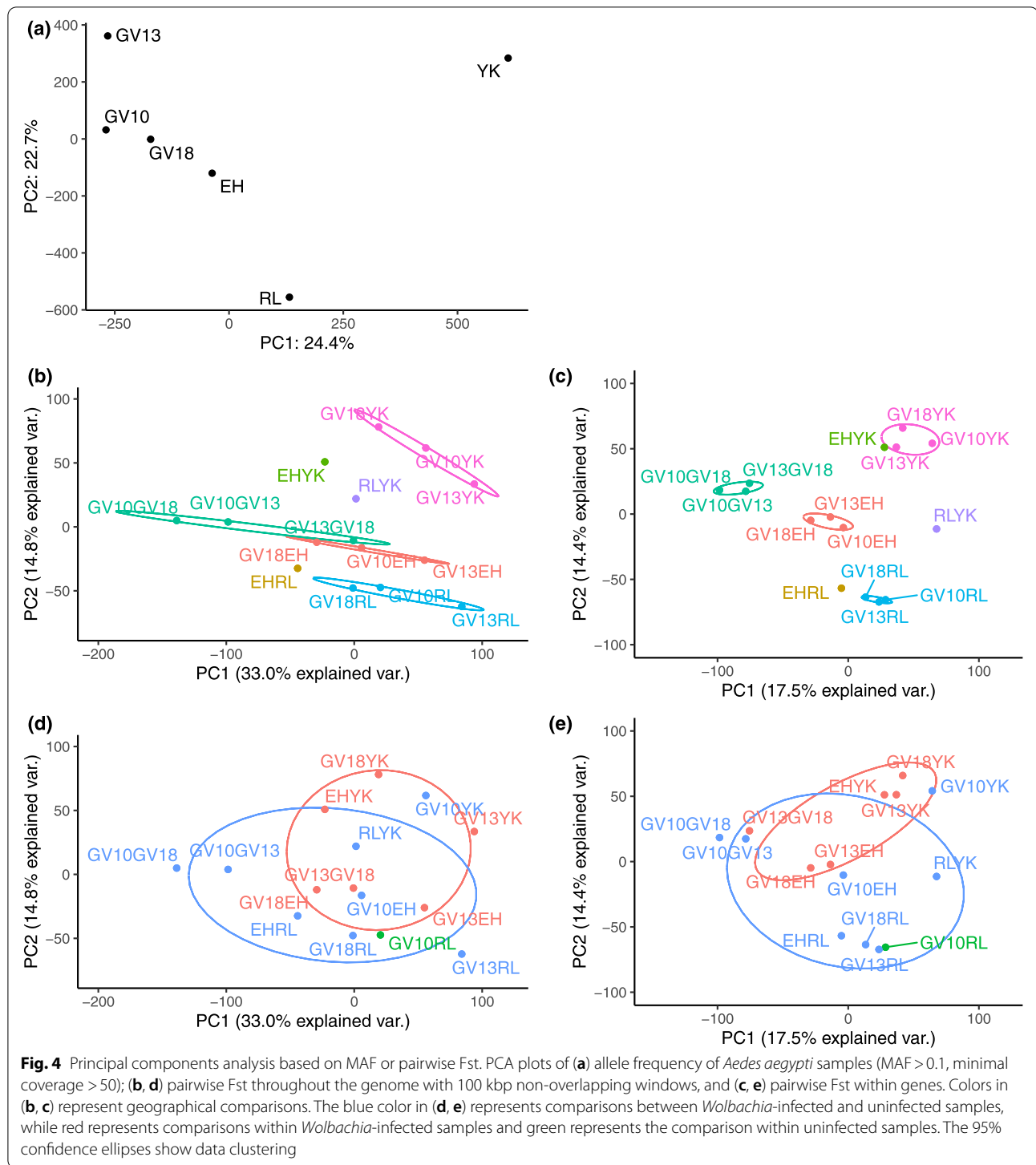
The introduction of linear relationships in the covariate model, however, can result in a high false positive rate from sampling noise, particularly when large environmental effects are involved and small number of populations are compared. We therefore used a second model to calculate  $X^T X$ , which was analogous to  $F_{st}$  [31, 32], 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%  $X^T X$  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 specific regions, with posterior probability at least five times greater than prior probability (> 3.7 in each Mbp), suggesting selection due to *Wolbachia* (Table 3). These SNPs were considered as “substantial” outliers associated with the *Wolbachia* infection.

We also used a stricter criterion: average  $BF^* > 10$  and falling into the intersection of top 5%  $X^T X$  values 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** Matrix of genetic distance (above diagonal) and geographic distance (below diagonal)

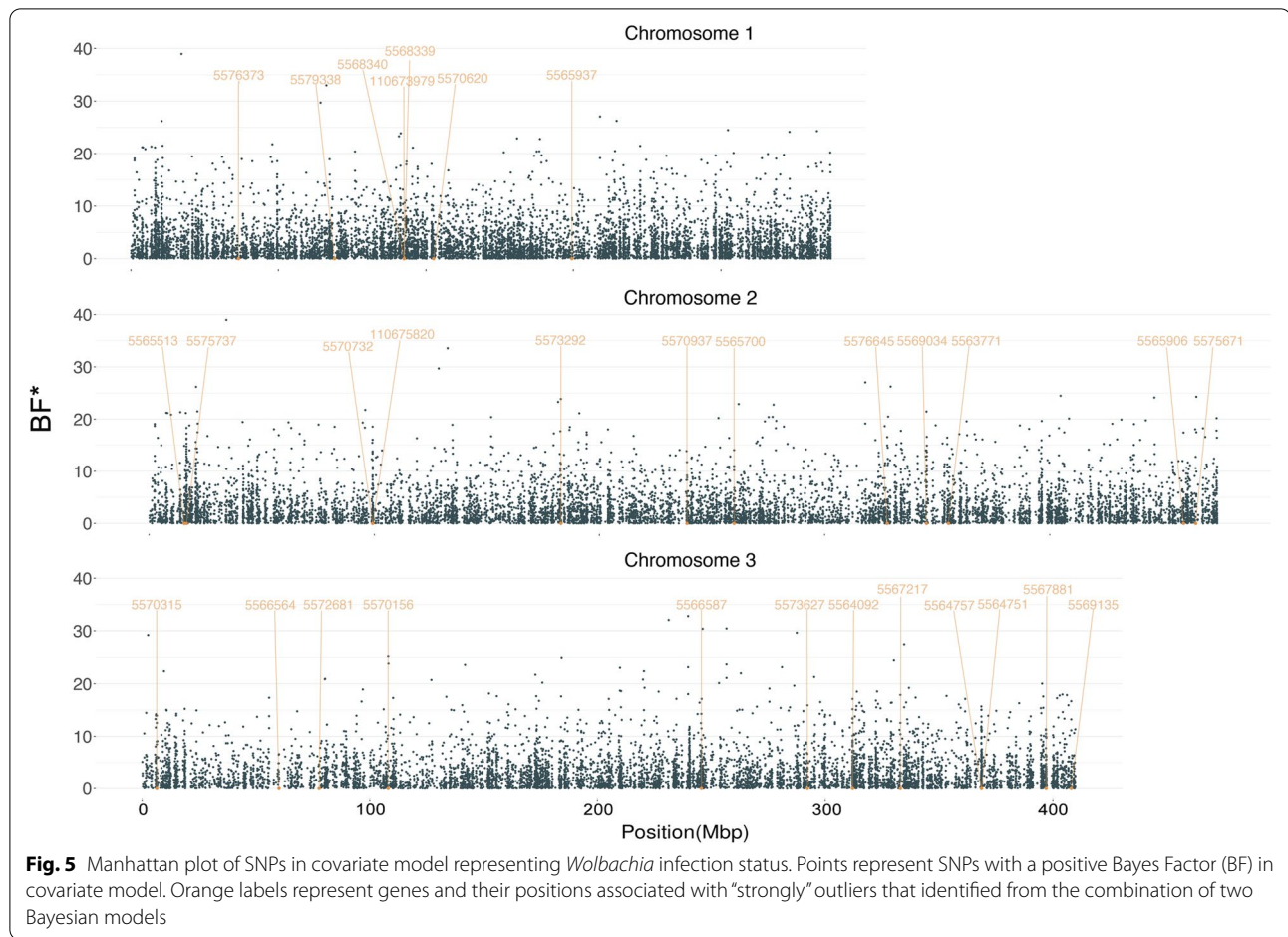
	GV10	GV13	GV18	EH	RL	YK
GV10	0	0.045	0.040	0.056	0.059	0.064
GV13	0	0	0.057	0.063	0.067	0.068
GV18	0	0	0	0.052	0.057	0.059
EH	3.243	3.243	3.243	0	0.051	0.054
RL	3.547	3.547	3.547	2.526	0	0.057
YK	3.694	3.694	3.694	2.821	2.773	0



#### Pathway analysis and gene ontology enrichment analysis

The “substantial” outliers were distributed within 187 genes (Additional file 2). When a region 50 kbp upstream and 50 kbp downstream around each outlier was considered, this number increased to 1436 genes

of interest (Additional file 3). We performed a pathway analysis of these 1436 genes through the KEGG database [33] and found eight pathways significantly involved (Table 4, Additional file 4), including pathways involved with development and immune response (MAPK



**Table 3** Genomic regions potentially under *Wolbachia* wMel selection

Chromosome	Position	Size (Mbp)	Number of SNPs
1	96,900,000-105,500,000	8.6	39
1	138,000,000-140,500,000	2.5	13
1	148,500,000-149,500,000	1.0	11
1	183,000,000-188,300,000	5.3	40
1	203,100,000-206,200,000	3.1	22
2	15,600,000-16,800,000	1.2	27
2	20,700,000-21,200,000	0.5	11
2	98,800,000-99,300,000	0.5	18
2	345,200,000-345,400,000	0.2	19
2	396,100,000-396,500,000	0.4	15
3	239,700,000-249,400,000	9.7	31
3	311,900,000-316,000,000	4.1	25
3	368,400,000-368,600,000	0.2	14

signaling pathway and Toll and Imd signaling pathway). We also identified homologues of these genes in *D. melanogaster* for the gene ontology (GO) enrichment analysis, 108 gene sets were significantly enriched in biological process (Additional file 5), 42 in cellular component (Additional file 6) and 22 in molecular function (Additional file 7). In general, we found that *Wolbachia* wMel may modulate genes with diverse functions such as cell development, interaction, cellular response, cellular transport, neurogenesis, lipid and glucide metabolism, behavior and immune response.

The “strong” outliers were distributed across 31 genes (Fig. 5, Additional file 1, Additional file 8). These included gene 5,564,751, which encodes cytochrome P450, which can be enriched in response to dengue virus infection in refractory mosquitoes [34] and is associated with insecticide resistance [35, 36]. Carbonic anhydrase, encoded by gene 5,565,700, has the function of balancing pH in mosquito midgut [37, 38]. Gene 5,569,135 encoding ecdysone protein E75, and gene 5,572,681 encoding lipophorin, have been highly expressed in females after blood feeding and are potentially involved in regulation

**Table 4** Significantly enriched pathways in KEGG database potentially involved with *Wolbachia* wMel infection (copyright permission 210,807) [33]

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

of oogenesis and vitellogenesis [39, 40]. Charged multivesicular body protein, encoded by gene 5,573,292, is associated with endosome formation and can influence mosquito immune response [41].

## Discussion

We show that *Ae. aegypti* populations in Cairns remain geographically distinct following releases of *Wolbachia* wMel, but also find some evidence suggesting evolutionary change in mosquito populations. When interpreting these results, it is important to consider the release process and which populations were targeted, 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]. In release areas around Cairns, Gordonvale is relatively isolated from other release locations. Although these populations are not genetically isolated based on microsatellite and EPIC markers [42], they do 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 GV10 population was a clear outlier.

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 hitchhikes along with the *Wolbachia* [12]. Also, while any linkage disequilibrium between the *Wolbachia* and nuclear DNA variants is expected to break down relatively quickly [43], 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 among the Gordonvale samples

before and after releases 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 differentiated at the gene level than at the genome level, which may be a consequence of a large proportion (47%) of TEs and satellites associated with non-coding regions [25, 26]. These TEs and satellites were masked in our reference genome [27], resulting in low coverage in sliding windows and were therefore filtered out in the gene analyses. This may explain why there was a high level of variation in the genome-wide scans of  $\pi$  and Tajima's D. For chromosome 1, which contains the sex-determining regions, pools of male and female individuals were aligned to the reference genome, which can also result in low coverage of windows and induce high levels of genomic variation. Genic regions will be easier to track into the future for further analysis.

When considering the impacts of *Wolbachia* on the *Ae. aegypti* genome, selection in response to local conditions and the impact of wMel on *Ae. aegypti* may only 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. When overlapping these results with previous Tajima's D analysis, we identified only a small proportion of genes potentially related to selective sweeps, which were mainly distributed on chromosome 2. These potential evolutionary impacts of *Wolbachia* wMel 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 Toll and Imd signaling pathways in KEGG analysis; these are important pathways in immune system [44, 45] and virus blocking processes [45–47]. Previous transcriptomic studies showed up-regulation of these pathways in both *wMel* and *wMel*-Pop-infected *Ae. aegypti* [48, 49]. Genes associated with virus blocking are mainly distributed on chromosome 1, in addition to genes associated with cytoskeleton, cell-cell adhesion and signal transduction [19]; these genes also showed up in our GO enrichment analysis. Other than virus blocking, caffeine metabolism was strongly impacted, which may impact hormone metabolism and detoxification when cytochrome P450 is involved [50, 51]. We also detected enriched pathways involved with development, such as insect hormone biosynthesis and the Hedgehog signaling pathways. In the GO enrichment analysis these were represented in cell growth, structure, recognition and behavior.

In the past decade, the *wMel* infection itself has not evolved in terms of either sequence composition or density since establishment in *Ae. aegypti* in northern Queensland, Australia [52]. 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 [53, 54], and may persist through ongoing selection favoring high viral blocking in *Ae. aegypti* populations [19]. However, the outlier loci detected here suggest that ongoing monitoring of phenotypic effects is warranted.

## Conclusions

*Wolbachia wMel*-infected *Ae. aegypti* mosquitoes have been released successfully in the field to help reduce the transmission of arboviruses, but interactions between *wMel* and *Ae. aegypti* could result in adaptation [55, 56], altering virus 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 wMel* 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 mosquito 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 in phenotypic assays of the longevity effects of *wMel*Pop in *Drosophila* [59].

## Methods

### Samples and study sites

*Aedes aegypti* samples were collected from four sites around Cairns (Fig. 1, Table 1) [60]. In Gordonvale, samples were collected three times: in the summer of 2010 (pre-release), as well as in 2013 and 2018 (2 and 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 [61]; 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 [62]. Samples from Gordonvale 2010 (GV10) and 2013 (GV13) were stored in absolute ethanol at the adult stage while samples collected from 2018 (GV18, EH, YK and RL) were stored in absolute ethanol 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, and using Roche High Pure™ 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 [63]. 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 1) 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 100bp paired read chemistry for GV10 and GV13, and 150bp 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 by requiring all reads to have all bases with a phred score >20 and read length >70bp. The reference genome AaegL5.0 [27] was indexed and reads were aligned using bowtie2 v. 2.3.4.3 with the --very-sensitive-local preset [64], with alignment rates ranging from 79.3 to 84.0%. Samtools v.1.9 was used with default parameters to sort, mark and

remove duplicates and generate pileup files requiring a minimal mapping quality of 20.

### Estimation of genetic variation

We investigated patterns of genetic variation within populations using PoPoolation v. 1.2.2 [65] with the genomic annotation file from the reference AegL5.0 which has TEs and satellites masked. We calculated Tajima's  $\pi$  (nucleotide diversity  $\pi$ ) for each population at 10 kbp non-overlapping windows and Tajima's D for each population at 10 kbp non-overlapping windows and at the gene level with a minimum coverage of 20. Windows with low coverage generated no values and were omitted before adjusting the shape of lines across each chromosome by a LOESS smooth curve [66]. Within each retained window, at least 60% of sites had sufficient coverage (Additional file 9), indicating the high quality of all windows used for downstream analysis. The value of Tajima's D was calculated from allele frequencies in selected regions and was used to detect directionality of selection. 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 genetic variation between populations, we used PoPoolation2 v. 1.201 [67] to obtain allele frequency differences for each SNP. The SNPs were filtered to have a coverage  $>50$  in all populations, ensuring that poorly sequenced areas were excluded. We also filtered to ensure an average MAF  $>0.1$  [68]. We also obtained pairwise Fst values for non-overlapping 100 kbp windows and for gene sets. The PCA was generated by the prcomp function and the ggbiplot package in R [69] 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) [70]; 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 IBD patterns among 2018 samples from the four locations by computing  $F_{st}^* = F_{st} / (1 - F_{st})$  [71] based on the average pairwise Fst values from 100 kbp 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 [72]. 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 [73], 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 [31] to identify outliers associated with *Wolbachia* infection. Firstly, we used a standard covariate model [31, 74], 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 relative evidence of a candidate SNP being under selection [75]. BF were exported in dB units ( $BF^* = 10 \times \log_{10}(BF)$ ); the association with environmental variance was considered as "substantial" when  $BF^* > 5$ , "strongly-selected" when  $BF^* > 10$  and "decisively-selected" when  $BF^* > 13$  following H Jeffreys [75]. 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 BF values are based on an Importance Sampling Approximation, which is unstable for single runs in particular when the number of population is small, so we averaged the BF values of three runs with different seeds from the random number generators following the suggestion in the manual of BayPass [31]. The introduction of linear relationships in this model, however, can lead to a high false positive rate from sampling noise, as the relationships of allele frequency among multiple populations are also influenced by other environmental factors, and are not independent from each other [32]. Based on this model, we generated plots for SNPs with average  $BF^* > 0$  to illustrate 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 as the main variable across time. An  $X^{TX}$  algorithm approach was used in this model, which was analogous to an Fst comparison [31, 32]. The  $X^{TX}$  value was used to identify selection pressure, with higher values suggesting positive selection and smaller values suggesting balancing selection [32]. 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. However, this model 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 the 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 frames to obtain a list of potentially important genes within a 100 kbp region around each SNP (50 kbp upstream and 50 kbp downstream) [76]. These genes were then searched through the KEGG pathway database [33] for interpretation. We then used BLAST+ (version 2.9.0) with the UniProtKB/Swiss-Prot database [77] to identify homologous proteins in *D. melanogaster*. Only best matches were retained and further filtered with an e-value cut off 1.0E-10 and >60% identity [78, 79]. Gene ontology (GO) enrichment analysis was then undertaken using the R packages clusterProfiler [80] and DOSE [81], with a false discovery rate cut-off of 0.05 after multiple hypothesis testing correction.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-08200-1>.

**Additional file 1.**

**Additional file 2.**

**Additional file 3.**

**Additional file 4.**

**Additional file 5.**

**Additional file 6.**

**Additional file 7.**

**Additional file 8.**

**Additional file 9.**

### Acknowledgements

We thank Gordana Rasic for her early contribution to the methodology for initial sequencing of the Gordonvale populations, and Moshe Jasper for his advice on data visualization. We also thank Scott Ritchie, Chris Paton and Mick Townsend from James Cook University for their support for sampling. 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).

### Authors' contributions

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

### Funding

This research was funded by the National Health and Medical Research Council (1132412, 1118640, [www.nhmrc.gov.au](http://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.

### Availability of data and materials

Aligned and merged sequencing data for *Aedes aegypti* used in this study are available from NCBI Sequence Read Archive (SRA), under the BioProject ID PRJNA776956. All data generated during this study are included within this article and its additional files.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Pest and Environmental Adaptation Research Group, Bio21 Institute and the School of BioSciences, The University of Melbourne, Parkville, Victoria, Australia. <sup>2</sup>Melbourne Bioinformatics, The University of Melbourne, Parkville, Victoria, Australia.

Received: 13 June 2021 Accepted: 15 November 2021

Published online: 14 December 2021

### References

- Nazni WA, Hoffmann AA, NoorAfizah A, Cheong YL, Mancini MV, Golding N, et al. Establishment of *Wolbachia* strain wAlbB in Malaysian populations of *Aedes aegypti* for dengue control. *Curr Biol*. 2019;29(24):4241–8.
- Garcia GA, Sylvestre G, Aguiar R, da Costa GB, Martins AJ, Lima JBP, et al. Matching the genetics of released and local *Aedes aegypti* populations is critical to assure *Wolbachia* invasion. *PLoS Negl Trop Dis*. 2019;13(1):e0007023.
- Ryan PA, Turley AP, Wilson G, Hurst TP, Retzki K, Brown-Kenyon J, Hodgson L, Kenny N, Cook H, Montgomery BL: Establishment of wMel *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.
- Van den Hurk AF, Hall-Mendelin S, Pyke AT, Frentiu FD, McElroy K, Day A, et al. Impact of *Wolbachia* on infection with chikungunya and yellow fever viruses in the mosquito vector *Aedes aegypti*. *PLoS Negl Trop Dis*. 2012;6(11):e1892.
- Aliota MT, Peinado SA, Velez ID, Osorio JE. The wMel strain of *Wolbachia* reduces transmission of Zika virus by *Aedes aegypti*. *Sci Rep*. 2016;6:28792.
- Walker T, Johnson PH, Moreira LA, Iturbe-Ormaetxe I, Frentiu FD, McMeniman CJ, et al. The wMel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. *Nature*. 2011;476(7361):450–3.
- Hoffmann AA, Montgomery B, Popovici J, Iturbe-Ormaetxe I, Johnson P, Muzzi F, et al. Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. *Nature*. 2011;476(7361):454–7.
- Turelli M, Hoffmann AA. Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations. *Genetics*. 1995;140(4):1319–38.
- Telschow A, Hammerstein P, Werren JH. Effects of *Wolbachia* on genetic divergence between populations: mainland-island model. *Integr Comp Biol*. 2002;42(2):340–51.
- 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.
- 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? *Mol Ecol*. 2011;20(17):3644–52.
- Hale LR, Hoffmann AA. Mitochondrial DNA polymorphism and cytoplasmic incompatibility in natural populations of *Drosophila simulans*. *Evolution*. 1990;44(5):1383–6.

13. Yeap HL, Rašić G, Endersby-Harshman NM, Lee SF, Arguni E, Le Nguyen H, et al. Mitochondrial DNA variants help monitor the dynamics of *Wolbachia* invasion into host populations. *Heredity* (Edinb). 2016;116(3):265–76.
14. Hoffmann AA, Iturbe-Ormaetxe I, Callahan AG, Phillips BL, Billington K, Axford JK, et al. Stability of the *wMel* *Wolbachia* infection following invasion into *Aedes aegypti* populations. *PLoS Negl Trop Dis*. 2014;8(9):e3115.
15. McMeniman CJ, Lane RV, Cass BN, Fong AWC, Sidhu M, Wang Y-F, et al. Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti*. *Science*. 2009;323(5910):141–4.
16. Ross PA, Wiwatanaratnabutr 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 *wMel* *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, et al. Selection on *Aedes aegypti* alters *Wolbachia*-mediated dengue virus blocking and fitness. *Nat Microbiol*. 2019;4(11):1832–9.
20. Ritchie SA, Townsend M, Paton CJ, Callahan AG, Hoffmann AA. Application of *wMelPop* *Wolbachia* strain to crash local populations of *Aedes aegypti*. *PLoS Negl Trop Dis*. 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 Biol*. 2007;5(5):e114.
22. Hornett EA, Charlat S, Duplouy AM, Davies N, Roderick GK, Wedell N, et al. 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):2059–68.
24. Ross PA, Turelli M, Hoffmann AA. Evolutionary ecology of *Wolbachia* releases for disease control. *Annu Rev Genet*. 2019;53:93–116.
25. Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, et al. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science*. 2007;316(5832):1718–23.
26. 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.
27. Matthews BJ, Dudchenko O, Kingan SB, Koren S, Antoshechkin I, Crawford JE, et al. Improved reference genome of *Aedes aegypti* informs arbovirus vector control. *Nature*. 2018;563(7732):501–7.
28. Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet*. 2008;9(5):397–405.
29. Lee Y, Schmidt H, Collier TC, Conner WR, Hanemaaijer MJ, Slatkin M, et al. Genome-wide divergence among invasive populations of *Aedes aegypti* in California. *BMC Genomics*. 2019;20(1):204.
30. 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.
31. Gautier M. Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*. 2015;201(4):1555–79.
32. Günther T, Coop G. Robust identification of local adaptation from allele frequencies. *Genetics*. 2013;195(1):205–20.
33. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
34. Behura SK, Gomez-Machorro C, Harker BW, deBruyn B, Lovin DD, Hemme RR, et al. Global cross-talk of genes of the mosquito *Aedes aegypti* in response to dengue virus infection. *PLoS Negl Trop Dis*. 2011;5(11):e1385.
35. 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 Negl Trop Dis*. 2012;6(6):e1692.
36. 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 Negl Trop Dis*. 2012;6(3):e1595.
37. Corena M, Seron TJ, Lehman HK, Ochrietor JD, Kohn A, Tu C, et al. Carbonic anhydrase in the midgut of larval *Aedes aegypti*: cloning, localization and inhibition. *J Exp Biol*. 2002;205(Pt 5):591–602.
38. del Pilar CM, VanEkeris L, Salazar MI, Bowers D, Fiedler MM, Silverman D, et al. Carbonic anhydrase in the adult mosquito midgut. *J Exp Biol*. 2005;208(Pt 17):3263–73.
39. Pierceall WE, Li C, Biran A, Miura K, Raikhel AS, Seagraves 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.
40. 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):733–8.
41. Bartholomay LC, Cho WL, Rocheleau TA, Boyle JP, Beck ET, Fuchs JF, 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):4114–26.
42. Endersby NM, Hoffmann AA, White VL, Lowenstein S, Ritchie S, Johnson PH, 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):1074–83.
43. Turelli M, Hoffmann AA, McKechnie SW. Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. *Genetics*. 1992;132(3):713–23.
44. Waterhouse RM, Kriventseva EV, Meister S, Xi Z, Alvarez KS, Bartholomay LC, et al. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science*. 2007;316(5832):1738–43.
45. Pan X, Pike A, Joshi D, Bian G, McFadden MJ, Lu P, et al. 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):277–88.
46. Xi Z, Ramirez JL, Dimopoulos G: The *Aedes aegypti* toll pathway controls dengue virus infection. *PLoS Pathog* 2008, 4(7):e1000098–e1000098.
47. Kambris Z, Blagborough AM, Pinto SB, Blagrove MSC, Godfray HCJ, Sinden RE, Sinkins SP: *Wolbachia* stimulates immune gene expression and inhibits plasmidium development in *Anopheles gambiae*. *PLoS Pathog* 2010, 6(10):e1001143–e1001143.
48. Kambris Z, Cook PE, Phuc HK, Sinkins SP. Immune activation by life-shortening *Wolbachia* and reduced filarial competence in mosquitoes. *Science*. 2009;326(5949):134–6.
49. 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.
50. Coelho A, Fraichard S, Le Goff G, Faure P, Artur Y, Ferveur JF, et al. Cytochrome P450-dependent metabolism of caffeine in *Drosophila melanogaster*. *PLoS One*. 2015;10(2):e0117328.
51. 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):5731–6.
52. 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.
53. 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 Negl Trop Dis* 2014, 8(2).
54. Carrington LB, Tran BCN, Le NTH, Luong TTH, Nguyen TT, Nguyen PT, 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):361–6.
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):1500–13.
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, et al. Rapid evolution of *Wolbachia* density in insecticide resistant *Culex pipiens*. *Heredity* (Edinb). 2010;104(1):15–9.
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):2648–57.
60. Kahle DJ, Wickham H. Ggmap: spatial visualization with ggplot2. *R J*. 2013;5(1):144.
61. Schmidt TL, Barton NH, Rašić G, Turley AP, Montgomery BL, Iturbe-Ormaetxe I, et al. Local introduction and heterogeneous spatial spread of dengue-suppressing *Wolbachia* through an urban population of *Aedes aegypti*. *PLoS Biol*. 2017;15(5):e2001894.
62. 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.
63. 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):4740–3.
64. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 2012;9(4):357–9.
65. 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.
66. Wickham H: ggplot2: elegant graphics for data analysis: Springer; 2016.
67. Kofler R, Pandey RV, Schlötterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*. 2011;27(24):3435–6.
68. Asif H, Alliey-Rodriguez N, Keedy S, Tamminga CA, Sweeney JA, Pearson G, et al. GWAS significance thresholds for deep phenotyping studies can depend upon minor allele frequencies and sample size. *Mol Psychiatry*. 2020.
69. Team RC. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019.
70. Hivert V, Leblois R, Petit EJ, Gautier M, Vitalis R. Measuring genetic differentiation from pool-seq data. *Genetics*. 2018;210(1):315–30.
71. Rousset F. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*. 1997;145(4):1219–28.
72. Medley KA, Jenkins DG, Hoffman EA. Human-aided and natural dispersal drive gene flow across the range of an invasive mosquito. *Mol Ecol*. 2015;24(2):284–95.
73. Diniz-Filho JAF, Soares TN, Lima JS, Dobrovolski R, Landeiro VL, de Campos Telles MP, et al. Mantel test in population genetics. *Genet Mol Biol*. 2013;36(4):475–85.
74. Coop G, Witonsky D, Di Rienzo A, Pritchard JK. Using environmental correlations to identify loci underlying local adaptation. *Genetics*. 2010;185(4):1411–23.
75. Jeffreys H. The theory of probability. Oxford: Oxford University Press; 1961.
76. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
77. 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.
78. Pearson WR: An introduction to sequence similarity ("homology") searching. *Curr Protoc Bioinformatics* 2013, Chapter 3:Unit3.1.
79. Tian W, Skolnick J. How well is enzyme function conserved as a function of pairwise sequence identity? *J Mol Biol*. 2003;333(4):863–82.
80. 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):284–7.
81. 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):608–9.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

