University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Schmidt, TL;Swan, T;Chung, J;Karl, S;Demok, S;Yang, Q;Field, MA;Muzari, MO;Ehlers, G;Brugh, M;Bellwood, R;Horne, P;Burkot, TR;Ritchie, S;Hoffmann, AA

Title:
Spatial population genomics of a recent mosquito invasion

Date:
2021-03-01

Citation:
Schmidt, T. L., Swan, T., Chung, J., Karl, S., Demok, S., Yang, Q., Field, M. A., Muzari, M. O., Ehlers, G., Brugh, M., Bellwood, R., Horne, P., Burkot, T. R., Ritchie, S. & Hoffmann, A. A. (2021). Spatial population genomics of a recent mosquito invasion. Molecular Ecology, 30 (5), pp.1174-1189. https://doi.org/10.1111/mec.15792.

Persistent Link:
https://hdl.handle.net/11343/298151

Check for updates

DR. THOMAS LUDOVIC SCHMIDT (Orcid ID : 0000-0003-4695-075X)

PROF. ARY HOFFMANN (Orcid ID : 0000-0001-9497-7645)

Article type : Original Article

Spatial population genomics of a recent mosquito invasion

Thomas L Schmidt*[1], T. Swan*[2,3], Jessica Chung[1,4], Stephan Karl[2,5], Samuel Demok[5], Qiong Yang[1], Matt A Field[2,6], Mutizwa Odwell Muzari[7], Gerhard Ehlers[7], Mathew Brugh[7], Rodney Bellwood[7], Peter Horne[7], Thomas R Burkot[2], Scott Ritchie[3,8], Ary A Hoffmann[1]

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

[2] Australian Institute of Tropical Health and Medicine, James Cook University, Cairns, QLD, Australia

[3] College of Public Health, Medical and Veterinary Sciences, James Cook University, Cairns, QLD, Australia

[4] Melbourne Bioinformatics, University of Melbourne, Parkville, Victoria, Australia

[5] Vector Borne Diseases Unit, Papua New Guinea Institute of Medical Research, Madang, Papua New Guinea

[6] John Curtin School of Medical Research, Australian National University, Canberra, ACT

[7] Medical Entomology, Tropical Public Health Services Cairns, Cairns and Hinterland Hospital & Health Services, Cairns, QLD, Australia

[8] Institute of Vector-Borne Disease, Monash University, Clayton, VIC, 3800, Australia

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/MEC.15792

This article is protected by copyright. All rights reserved

Author Manuscript

* equal contributions

Contact: tom.schmidt@unimelb.edu.au

Running title: Population genomics of a recent invasion

Abstract

Population genomic approaches can characterise dispersal across a single generation through to many generations in the past, bridging the gap between individual movement and intergenerational gene flow. These approaches are particularly useful when investigating dispersal in recently altered systems, where they provide a way of inferring long-distance dispersal between newly established populations and their interactions with existing populations. Human-mediated biological invasions represent such altered systems which can be investigated with appropriate study designs and analyses. Here we apply temporally-restricted sampling and a range of population genomic approaches to investigate dispersal in a 2004 invasion of *Aedes albopictus* (the Asian tiger mosquito) in the Torres Strait Islands (TSI) of Australia. We sampled mosquitoes from 13 TSI villages simultaneously and genotyped 373 mosquitoes at genome-wide single nucleotide polymorphisms (SNPs): 331 from the TSI, 36 from Papua New Guinea (PNG), and 4 incursive mosquitoes detected in uninvaded regions. Within villages, spatial genetic structure varied substantially but overall displayed isolation by distance and a neighbourhood size of 232–577. Close kin dyads revealed recent movement between islands 31–203 km apart, and deep learning inferences showed incursive *Ae. albopictus* had travelled to uninvaded regions from both adjacent and non-adjacent islands. Private alleles and a coancestry matrix indicated direct gene flow from PNG into nearby islands. Outlier analyses also detected four linked alleles introgressed from PNG, with the alleles surrounding 12 resistance-associated cytochrome P450 genes. By treating dispersal as both an intergenerational process and a set of discrete events, we describe a highly interconnected invasive system.

This article is protected by copyright. All rights reserved

Author Manuscript

Keywords: **dispersal, *Aedes albopictus*, adaptive introgression, genetic invasion, kinship, biological invasion**

Introduction

Population genetics has traditionally treated dispersal as an intergenerational process, but one that is nevertheless derived from the movement of individual organisms within each generation (Wright, 1943). Intergenerational dispersal describes how organisms distributed across geographical space are connected through time via a spatial pedigree (Bradburd & Ralph, 2019), and can be summarised at a population level by the mean distance between parent and offspring. Population genomic approaches applied to wild populations increasingly provide the power needed to detect dispersal across fine temporal scales down to single generations, which can reveal movement patterns at correspondingly fine spatial scales (Combs, Puckett, Richardson, Mims, & Munshi-South, 2018; Jasper, Schmidt, Ahmad, Sinkins, & Hoffmann, 2019; Trense et al., 2020). Spatial genomic studies conducted across a restricted temporal range may also help shed new light on adaptive processes such as the spread of advantageous alleles through wild populations (Endersby-Harshman et al., 2020; Fitzpatrick et al., 2010; Pélissié, Crossley, Cohen, & Schoville, 2018) and local adaptation in microgeographically structured environments (Yadav, Stow, & Dudaniec, 2020).

This article is protected by copyright. All rights reserved

Author Manuscript

Individuals connected through recent generations of the spatial pedigree will be close kin, with full-sibs and half-sibs separated by a single generation and first cousins by two. Close kin dyads can be identified using genomics, and the spatial distribution of kin has been used in recent studies to detect dispersal over fine temporal scales (Combs et al., 2018; Escoda, Fernández-González, & Castresana, 2019; Escoda, González-Esteban, Gómez, & Castresana, 2017; Fountain et al., 2018; Jasper et al., 2019; Schmidt, Filipović, Hoffmann, & Rašić, 2018; Trense et al., 2020). Dispersal inferences from close kin treat dispersal as a set of discrete events reflecting specific acts of individual movement, making them particularly valuable for investigating dispersal through regions of genetic similarity, such as where populations have only recently become isolated (Escoda et al., 2019, 2017) or where a population has been sampled continuously across a range (Combs et al., 2018; Jasper et al., 2019; Schmidt et al., 2018; Trense et al., 2020). As kinship-based approaches assess recent movement, they are ideally applied to systems which have been subject to recent change, such as biological invasions or threatened species in disturbed habitats.

While the past 1-2 generations of dispersal represent very recent movement, population genomics has also been used to detect intragenerational dispersal, which has traditionally been investigated by directly observing the movement of individuals (Harrington et al., 2005; Schultz & Crone, 2001). Examples of this approach include tracing individual invasive species incursions ("incursives") to their population of origin (M. Z. Chen et al., 2020; Schmidt, Chung, van Rooyen, et al., 2020; Schmidt et al., 2019). Although such studies require sampling across broad scales they are a useful way of detecting long-distance dispersal, which is difficult to investigate but can have important evolutionary consequences (Gillespie et al., 2012; Waters, Fraser, & Hewitt, 2013). Note that when incursive individuals are intercepted, there is no 'dispersal' in the intergenerational sense, but these movements are nevertheless important components of life-histories and food webs (Howard, 1960).

Recent human-mediated biological invasions are complex spatial processes that require careful investigation. In recent invasions, high regional coancestry throughout the invaded region will

This article is protected by copyright. All rights reserved

Author Manuscript

make the analysis of genetic structure alone insufficient for evaluating movement (Fitzpatrick, Fordyce, Niemiller, & Reynolds, 2012). Additionally, invasive species, defined here as any taxon which rapidly spreads (following Cristescu (2015)), frequently exhibit stratified dispersal, in which short-range active movement by the organism and long-range passive transportation by humans operate together across a range of spatial scales within a generation (Hengeveld, 1989; Sharov & Liebhold, 1998). Long-range dispersal can be traced within invaded regions or from distant origins (Schmidt, Chung, Honnen, Weeks, & Hoffmann, 2020; Sherpa et al., 2019), and can operate alongside short-range dispersal to spread adaptive alleles into and through invasive populations (Endersby-Harshman et al., 2020; Pélissié et al., 2018). Applying intragenerational and kinship-based dispersal methods to recent invasions may help detect recent movement across fine and broad spatial scales, with dispersal between invaded regions revealed through close kin and intragenerational incursions into uninvaded regions traced to their source.

Here we use spatial population genomics to investigate an invasion of *Aedes albopictus* (Asian tiger mosquito) in the Torres Strait Islands (TSI), Australia, sampled ~14 years after colonisation. The Torres Strait lies between the northernmost point of the Australian mainland (Cape York) and Papua New Guinea (PNG) (Fig 1 inset). The region contains over 100 islands, of which 18 are inhabited communities (http://www.tsra.gov.au/the-torres-strait/community-profiles#TS%20Communities) and which we refer to here as 'villages' to distinguish from ecological communities. *Aedes albopictus* was first detected in the TSI in 2004 (Ritchie et al., 2006), and an insecticide-focussed elimination program began shortly thereafter. Logistical difficulties and recurrent reinvasion risks led to the abandonment of this program in 2008 in favour of a containment strategy, in which a *cordon sanitaire* was imposed to regulate the movement of people and goods between the invaded islands of the Torres Strait and the uninvaded 'inner' islands of Ngurapai (Horn) and Waiben (Thursday) (Muzari et al., 2017; van den Hurk et al., 2016). Ngurapai and Waiben serve as commercial hubs connecting the TSI and the Australian mainland, and the *cordon sanitaire* approach has likely helped prevent incursion of *Ae. albopictus* onto the mainland of Australia through the Torres Strait route though it continues to vector dengue outbreaks in the TSI (Muzari et al., 2017). *Aedes albopictus*

This article is protected by copyright. All rights reserved

Author Manuscript

incursions continue to be detected at ports on the Australian mainland and have been traced to locations in East and Southeast Asia that have strong trade links to Australia (Schmidt, Chung, van Rooyen, et al., 2020). Previous investigations of TSI *Ae. albopictus* using mtDNA and microsatellites indicated high coancestry with Indonesian *Ae. albopictus* (Beebe et al., 2013; Maynard et al., 2017). Additional observations of spatially and temporally variable genetic structure within the TSI could reflect high regional gene flow following local founder events (Maynard et al., 2017).

While Indonesia is the likely source of the initial invasion into the TSI, over 5,000 boat journeys are made each year between TSI and PNG (Mcfarlane, 1998), indicating the potential for gene flow across this border. When gene flow takes place after an initial invasion it can operate as a 'genetic invasion' if it interferes with control strategies, such as through introgression of alleles conferring insecticide resistance (Endersby-Harshman et al., 2020; Riveron et al., 2013; Schmidt et al., 2019). Although the TSI was likely invaded from Indonesia via a 'stepping-stone' in the Southern Fly region of PNG (Beebe et al., 2013), the absence of *Ae. albopictus* in Southern Fly in the late 1990s (Johansen et al., 2000) and its present genetic similarity to the TSI (Maynard et al., 2017) indicates that the invasion of the TSI and Southern Fly regions happened contemporaneously from Indonesia. PNG *Ae. albopictus* from outside the Southern Fly region are genetically distinct from Indonesian *Ae. albopictus* (Maynard et al., 2017) and gene flow from this PNG background should be readily detectable.

This investigation covers a range of spatial scales from 10s of metres to 1000s of kilometres, and temporal scales from a single generation to ~100 generations in the past. Within the TSI, where coancestry is high, we focus on dispersal in the immediate past, covering both short-range active flight within islands and long-range passive transportation between islands, including incursions past the *cordon sanitaire*. At broader spatial scales, we use a panel of differentiated genotypes from other locations including nearby PNG to investigate earlier patterns of gene flow into the TSI. Our specific aims are: (i) to investigate the spatial structure of passive dispersal, and whether it occurs only between nearby locations or distant locations as well; (ii)

This article is protected by copyright. All rights reserved

Author Manuscript

to ascertain whether new genetic structure among islands has developed since previous investigation (Maynard et al., 2017); and (iii) to determine whether gene flow has occurred from PNG into the TSI, including the spread of putatively adaptive alleles. Through this, we show how stratified dispersal operates in this system and how long-range dispersal can assist the rapid invasion of new regions and the spread of advantageous alleles through established populations. The rapid spread of *Ae. albopictus* and its capacity for long-distance invasion make studies of its dispersal globally relevant (Goubert et al., 2017; Schmidt, Chung, Honnen, et al., 2020; Sherpa et al., 2019).

Materials and methods

Sample acquisition and genotyping

We analysed *Ae. albopictus* single nucleotide polymorphism (SNP) data from multiple sources. These included new mosquito samples collected from the Torres Strait Islands (TSI) and Papua New Guinea (PNG) as well as sequence data collected previously from across the Indo-Pacific region. We use the alphabetical key (A-L) in Fig 1 in all references to TSI villages.

Sampling in the invaded regions of the TSI took place at two time points. The first involved cross-sectional collections from 13 villages on 12 islands sampled between 2018-04-24 and 2018-05-04 (Fig 1). On Moa Island, two villages were sampled (B:St Pauls and C:Kubin). By restricting the temporal range of sampling, it is possible to undertake investigations within and between villages that require data at a fine temporal scale (i.e. within 1–2 generations), such as kinship analysis. Within each village, researchers selected and georeferenced sampling points where they collected adult *Ae. albopictus* with sweep-nets. Sampling points were separated by at least 100 m where possible (maximum 3,297 m), and each village had 15–20 sampling points

This article is protected by copyright. All rights reserved

Author Manuscript

except Dauan which had 7 (Fig 1: red crossed square). *Aedes albopictus* were preserved in 100% ethanol prior to DNA extraction. A second TSI sampling event took place on I:Masig between 2019-03-28 and 2019-04-08 at the same sampling points, serving as a temporal replicate. Village maps and sampling point locations are detailed in Figures S1-S13.

In February 2019, we detected *Ae*. *albopictus* incursions on the island of Ngurapai, south of the *cordon sanitaire* (see Fig 2). Four incursive *Ae. albopictus* adults were detected between 2019-02-02 and 2019-02-09 during routine wet season surveillance (Figure S14) and the samples were preserved for DNA extraction. Similar surveys, conducted with sweep nets in December 2018, March 2019 and May 2019 did not detect any *Ae. albopictus* on the island.

A final set of samples was collected from PNG between 2018-08-02 and 2018-09-13 using ovitraps deployed at several urban locations around the cities of Port Moresby (south coast) and Madang (north coast). These were aggregated to produce one Port Moresby sample and one Madang sample, with a maximum of two mosquitoes taken from each individual ovitrap to reduce the likelihood of sampling close kin.

Samples were genotyped using a pipeline described previously (Schmidt, Chung, van Rooyen, et al., 2020) and described in full in Text S1. Briefly, we used the double digest restriction site-associated (ddRAD) sequencing protocol developed for *Ae. aegypti* (Rašić, Filipović, Weeks, & Hoffmann, 2014) and validated in *Ae. albopictus* (Schmidt et al., 2017). Sequence reads were processed in Stacks v2.41 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013), with Bowtie2 v2.3.5.1 (Langmead & Salzberg, 2012) used to align reads to the AalbF2 genome assembly (Palatini et al., 2020).

Sequence data were generated for 371 *Ae. albopictus*: 301 from the cross-sectional collections on the invaded islands, 30 from the temporal replicate on I:Masig, 4 from the incursions past the *cordon sanitaire*, and 18 from each of Madang and Port Moresby in PNG. We also included sequence data from previous work (Schmidt, Chung, Honnen, et al., 2020; Schmidt, Chung, van Rooyen, et al., 2020). This included sequences from Indonesia (Bali, Bandung and Jakarta) and

This article is protected by copyright. All rights reserved

Author Manuscript

Timor-Leste, the proposed approximate origin of the initial TSI invasion (Beebe et al., 2013; Maynard et al., 2017). We also added sequences from locations linked to Australian *Ae. albopictus* incursions: China (Guangzhou), Japan, Singapore and Taiwan (Schmidt, Chung, van Rooyen, et al., 2020). Finally, we included two Pacific Island locations from East of PNG: Fiji and Vanuatu. These data were all reprocessed through the same bioinformatics pipeline.

Datasets and filtering

Genotypes were filtered with the Stacks v2.41 program Populations (Catchen et al., 2013), VCFtools v0.1.16 (Danecek et al., 2011), and Beagle v4.1 (Browning & Browning, 2016). Populations produced VCF files containing SNPs called in ≥ 0.75 of the mosquitoes from each village, location or timepoint; VCFtools retained biallelic SNPs with minor allele count ≥ 3 (Linck & Battey, 2019), depth of coverage ≥ 5, and genotyped in ≥ 80% of total mosquitoes; and Beagle imputed and phased genotypes in 50,000 bp sliding windows with 3,000 bp overlap.

We first used Populations and VCFtools to detect close kin dyads (1^{st} and 2^{nd} order) within each village. We used the *relatedness2* command in VCFtools to detect first-order (parental or full-sib) and second-order (grandparental or half-sib) kin dyads, analysing each village separately. This command uses the KING (Manichaikul et al., 2010) method of generating kinship coefficients for pairwise relationship inference; accordingly, we used kinship coefficient cut-offs of > 0.177 for first-order kin and > 0.088 for second order kin. Close kin dyads identified at this step were used to help estimate kinship categories of dyads across villages (see 'Close kin across villages' below).

For analysis within the TSI, we were cautious to avoid filtering bias caused by close kin and uneven sample sizes (Puechmaille, 2016). Accordingly, we produced datasets of n = 22 for each village, with 22 the maximum number of genotypes that ensured equal sample sizes after

This article is protected by copyright. All rights reserved

Author Manuscript

removing close kin within and across villages (see Results). Kin were removed in order of missing data, other genotypes were removed to maximise the geographical distribution of genotypes, then by missing data. Datasets were imputed and phased with Beagle and hereafter called the "Village22" datasets, each containing between 7,877–9,873 SNPs.

We also produced a dataset for the TSI region called "Region22" that contained all individuals from the Village22 datasets, containing 26,085 SNPs. Tracing of incursions and gene flow from PNG used this dataset where specified, with the incursive or PNG genotypes included. For analyses using the additional Indo-Pacific genotypes, we included all non-incursive individuals from the TSI including Dauan, with the temporal replicate from I:Masig excluded and with the dataset left unimputed where indicated. This dataset contained 495 mosquitoes genotyped at 26,970 SNPs.

Population processes within the Torres Strait

TSI genetic structure

We investigated genetic structure in the TSI using sparse non-negative matrix factorisation (sNMF) in the R package "LEA" (Frichot & François, 2015), run on the Region22 dataset. This analysis estimates individual ancestry coefficients, assuming that individual genotypes are produced from the admixture of K ancestral lineages. Previous research using microsatellites to analyse samples from 2007–2015 found two main genetic lineages in the TSI (Maynard et al., 2017). Accordingly, we ran 50 iterations of the sNMF with $2 \leq K \leq 4$ but used $K = 2$ for visualisations to provide a direct comparison with this work.

Author Manuscript

This article is protected by copyright. All rights reserved

We used Populations to calculate heterozygosity (H_O) and nucleotide diversity (π) at variant sites and pairwise F_{ST} between all village samples. We used VCFtools to calculate Tajima's *D* in 10 Mbp windows, including the temporal replicate from I:Masig to reveal any changes over the 12 month period. We used the R package "hierfstat" (Goudet, 2005) to estimate global F_{ST} and the population differentiation statistic D_{est} (Jost, 2008) among the 2018 samples.

Dispersal within villages

Within TSI villages (< 3.5 km), dispersal is mostly by active flight through continuous habitat and can be summarised by the parameter σ, which makes up the dispersal variance component of neighbourhood size (N_W: Wright, 1946). N_W can be thought of as the effective number of *Ae. albopictus* that make up a mosquito's breeding 'neighbourhood' when isolation by distance is operating. The equation, $N_W = 4\pi.\sigma^2.d$, provides a link between spatial patterns of genetic differentiation and local demographic processes, connecting intergenerational dispersal (σ^2) to the effective density of breeding adults within the dispersal area (d).

We used the Village22 datasets to investigate isolation by distance within each village. Mantel tests (function mantel.randtest) ran in the R package "ade4" (Dray & Dufour, 2007) analysed matrices of individual linear genetic distance and the natural logarithm of Haversine geographic distance, using 9,999 permutations and Bonferroni correction to assess statistical significance. Rousset's *a* (Rousset, 2000) provided genetic distance, calculated in SPAGeDI (Hardy & Vekemans, 2002).

We estimated neighbourhood size (N_W: Wright, 1946) within all villages using the inverse of the regression slope of pairwise genetic distance (Rousset's *a*) against the natural logarithm of geographical distance (Rousset, 2000). As the linearity of this relationship may break down at distances within the dispersal estimate σ, we omitted pairs < 50 m apart. The temporal replicate

This article is protected by copyright. All rights reserved

Author Manuscript

from I:Masig was also omitted. The pairwise data used for Mantel tests were concatenated to run a single linear regression on the 5325 within-village dyads.

Close kin across villages

Having identified close kin dyads within villages, we applied a conservative process for detecting kin across villages, which represent passive transportation by humans. First, we generated kinship coefficients using PC-Relate (Conomos, Reiner, Weir, & Thornton, 2016), which controls for genetic structure by conditioning the data with principal components (PCs). We ran PC-Relate on the dataset containing all non-incursive individuals, first pruning SNPs by linkage disequilibrium (R package "SNPRelate" (Zheng et al., 2012); using the snpgdsLDpruning function with arguments, method = "composite" and ld.threshold = 0.2).

We generated kinship coefficients for all dyads within and across villages following three different conditioning treatments: 3 PCs, 5 PCs, and 10 PCs. As a lower bound for determining which dyads across villages could be considered to have a particular category of relatedness, we used the lowest kinship coefficient observed among all close kin dyads within villages (see Datasets and filtering).

For each dyad found across two villages, one individual (Sib-1) or its recent relative will have dispersed to a new environment (Village-1) while the other (Sib-2) has stayed at the origin (Village-2). Thus, Sib-1 and Sib-2 should both have higher coancestry with individuals from Village-2 than those from Village-1, which should be reflected in higher average kinship with Village-2. We estimated the direction of dispersal by comparing kinship coefficients of Sib-1 and Sib-2 with each village, which allowed us to determine the origin (Village-2) and the destination (Village-1). These differences were evaluated with t-tests.

Author Manuscript

This article is protected by copyright. All rights reserved

Incursions past the *cordon sanitaire*

To estimate source locations of the four incursive *Ae. albopictus*, we used two complementary methods in the programs assignPOP (K.-Y. Chen et al., 2018) and Locator (Battey, Ralph, & Kern, 2020). assignPOP treats each village as a population, and generates assignment probabilities to each hypothetical population using a Monte Carlo assignment test with a support vector machine predictive model. Locator makes no population-based assumptions, but uses geolocations of individual genotypes to infer the spatial origin of each incursive using deep learning. These analyses used the Region22 dataset, omitting the temporal replicate from I:Masig.

Following Schmidt et al. (2019), we used the assignPOP function assign.X to generate posterior probabilities of assignment from which we calculated 'relative probabilities' of assignment, defined as the highest posterior probability divided by the second highest. Relative probability > 3 has previously been used as a cut-off for *Ae. albopictus* assignment at broad scales (Schmidt, Chung, van Rooyen, et al., 2020).

To run Locator we first pruned SNPs by linkage disequilibrium using the snpgdsLDpruning function in the R package "SNPRelate" (arguments, method = composite and ld.threshold = 0.2). This dataset provided a point estimate of each incursive's original location, and 1000 bootstrap subsamples were run to provide confidence around these estimates. While Locator has options for analysing SNPs in windows, our ddRADseq data were too sparse and the reference assembly insufficient for this.

Before running final assignments, we first confirmed that each of the four incursives had an origin in the TSI through an initial run of assignPOP using the dataset containing all non-incursive Indo-Pacific genotypes. These results showed strong support for each incursive mosquito having an origin in the TSI, with aggregated posterior assignment probabilities of 0.89–0.97 to the TSI and 0.11–0.03 to non-TSI locations (Table S1).

This article is protected by copyright. All rights reserved

Author Manuscript

International gene flow into the Torres Strait

Genome-wide gene flow

We used fineRADstructure (Malinsky, Trucchi, Lawson, & Falush, 2018) to investigate possible coancestry between TSI *Ae. albopictus* and those from locations other than Indonesia or Timor-Leste. We ran fineRADstructure with default settings on the dataset containing all non-incursive Indo-Pacific genotypes, with the temporal replicate of I:Masig removed.

This analysis also helped clarify genetic relationships among genotypes in the broader region. To assist with the visualisation of these relationships, we ran a PCA-UMAP analysis (Diaz-Papkovich, Anderson-Trocmé, Ben-Eghan, & Gravel, 2019) on the same dataset. PCA-UMAP used the first 4 PCs, projected in two dimensions via UMAP using 50 neighbours and a 0.5 minimum distance.

Using the Region22 dataset, we calculated the number of private alleles in each TSI village with Populations, excluding the temporal replicate from I:Masig. This was then recalculated with the two PNG samples (Port Moresby and Madang) included in filtering. By comparing the number of private alleles in each village with and without the PNG samples, we assessed gene flow between each village and PNG. Specifically, if fewer private alleles were recorded for a village after the PNG samples were included in filtering, the "missing" alleles will be identical by state and likely identical by descent in both PNG and that village. As these alleles are restricted to specific villages, they are evidence of gene flow from PNG into each village directly.

Adaptive introgression

While ddRAD data are typically too sparse to detect alleles under selection, they can be suitable for detecting selective sweeps as shown by previous work on *Ae. aegypti* (Endersby-Harshman et al., 2020). We looked for signs that advantageous alleles had spread into the TSI using a

This article is protected by copyright. All rights reserved

Author Manuscript

three-step process: identifying genomic regions of interest, where multiple SNPs had irregular structure consistent with linked selection; identifying genes within these genomic regions of potential adaptive importance; and identifying the geographical origin of alleles at these SNPs.

We used the R packages *pcadapt* (Luu, Bazin, & Blum, 2017) and LEA (Frichot & François, 2015: function sNMF) to detect SNPs with irregular structure. Analyses were run on the Region22 dataset. We ran *pcadapt* with a minimum allele frequency of 0.025 (≥ 15 allele copies present), and obtained P-values from Mahalanobis distances. Genomic inflation factors for all runs were < 1.5. We ran sNMF using default parameters and 50 repetitions, and selected outliers by F_{ST}. For each analysis, P-values were transformed using a Bonferroni correction, and we used a q-value cut-off of 0.001. Only SNPs detected as outliers by both *pcadapt* and sNMF were considered. As genotype imputation methods can introduce bias when imputing rare alleles (Shi et al., 2018), we repeated the above methods without imputation in Beagle.

Within TSI *Ae. albopictus*, local demographic changes such as founder effects may have eliminated genetic diversity in particular regions of the genome, which a genome scan may incorrectly read as the result of a selective sweep (Hoban et al., 2016). For this reason, we restricted our investigation to outlier regions that satisfied three conditions: (i) the region contained multiple outlier SNPs spread across multiple RADtags within 1 Mbp; (ii) the region contained one or more genes with products of plausible adaptive importance, such as insecticide resistance; and (iii) the alleles at the specific outlier SNPs within the region were not found in Indonesia (putative source of the TSI invasion) but were found in other populations outside of the TSI. For (ii) we used Integrative Genomics Viewer (Robinson et al., 2011) to visually explore the *Ae. albopictus* reference assembly. For (iii), we used the dataset containing all non-incursive Indo-Pacific genotypes (N = 525), without imputation.

Author Manuscript

This article is protected by copyright. All rights reserved

Results

Population processes within the Torres Strait

TSI genetic structure

Analysis of genetic structure with sNMF provided a picture of genetic patterns across the TSI for K = 2 (Fig 1). Despite the higher resolution of SNP markers and the additional ~3-11 years elapsed since sampling, there was no clearer spatial genetic structure within and between villages at K = 2 than in the previous microsatellite study (c.f. Maynard et al., 2017). Both ancestral lineages were detected in every village, though 48 individuals were assessed as not admixed (Fig 1). Genetic admixture among individuals was highly variable in some villages, particularly in D:Badu, G:Warraber, I:Masig and L:Mer. Despite this, some geographical patterns were evident from the K = 2 analysis, particularly the genetic similarity of the central islands (F:Iama, G:Warraber, H:Poruma) and those of Moa Island (B:St Pauls and C:Kubin). These patterns are increasingly pronounced at K = 3 (Figure S16), though the high admixture within villages (Figures 1,S15,S16,S17) and low pairwise F_{ST} among villages (Table S2) indicate that genetic structure is weak within the TSI. Genetic similarity among the central islands has previously been noted (Maynard et al., 2017), and points to temporal consistency in genetic structure.

Mean global F_{ST} among villages in 2018 was 0.039 (SE = 0.000) while differentiation (D_{est}) was 0.006 (SE = 0.000). Pairwise F_{ST} estimates between villages (Table S2) indicate that I:Masig had higher average pairwise F_{ST} in 2019 ($\bar{x}$ = 0.032, SE = 0.001) than in 2018 ($\bar{x}$ = 0.029, SE = 0.001), an average increase in F_{ST} of 8.9% over the 12 months, perhaps reflecting the development of additional genetic structure across time. Tajima's D can be difficult to interpret, but the

Author Manuscript

This article is protected by copyright. All rights reserved

399 consistently negative mean values across all villages accord with past population expansions
400 following colonisation (Gattepaille, Jakobsson, & Blum, 2013).

401

402 ***Dispersal within villages***

403 Table 1 lists mean pairwise genetic distance (Rousset's *a*) for dyads sampled at < 50 m and > 500
404 m separation within each village. Mosquitoes sampled at > 500 m separation did not have
405 consistently higher genetic distances than those sampled at < 50 m separation. Likewise, 6 of
406 the 13 Mantel tests showed a negative relationship between genetic and geographic distance
407 within villages (Figures S18-S30). There was no statistically significant isolation by distance
408 within any of the villages (all Bonferroni-corrected P-values > 0.05).

409 When within-village dyads were concatenated to estimate neighbourhood size (N_W), isolation by
410 distance was revealed by the positive relationship between genetic distance (Rousset's *a*) and
411 the natural logarithm of geographical distance (geodist): $a = 0.003025 \times \text{geodist} + 0.1878$, with
412 coefficient SE = 0.00129. N_W within villages was estimated at $\frac{1}{0.003025} = 331$ mosquitoes (95%
413 C.I. 232–577). Despite high variability among villages, these results indicate that TSI *Ae.*
414 *albopictus* have some spatial structure at scales of 100s to 1000s of metres.

415

416 ***Close kin within villages***

417 Three of the 2772 dyads within villages had putative first-order relatedness, and none had
418 second-order relatedness (Table 2). Two of the three dyads were collected at the same sampling
419 point, the other at points 170 m apart. This low incidence of close kin compared with other
420 studies (e.g. Jasper et al., 2019) was likely due to the 100 m separation between most sampling
421 points and suggests mostly local movement within villages.

422

This article is protected by copyright. All rights reserved

Author Manuscript

Close kin across villages

Within villages, the dyad in E:Mabuiag had the lowest PC-Relate kinship coefficient of the three close kin dyads (0.32 with 3 PCs; Table 2), so this was used as the cut-off for designating close kin across villages. Five of the 31,944 dyads across villages had higher scores than this, indicating putative first-order relatedness. The three highest kinship coefficients for 'unrelated' dyads were 0.27, 0.23, and 0.16. When 5 or 10 PCs were included, the rank order of kinship coefficients fluctuated but all remained consistently above the within-village cut-off and above the coefficients of unrelated dyads.

Four of the putative close kin dyads were across nearby villages while the fifth was between E:Mabuiag and L:Mer, ~200 km apart (Figure 1, Table 2). For each dyad the direction of dispersal was apparent, demonstrated by both individuals in the dyad having higher mean kinship coefficients with Village-2 (the origin) than Village-1 (the destination). Results of t-tests assessing dispersal direction were all strongly statistically significant (all $t > 8.6$, all $P < 0.001$).

Incursions past the *cordon sanitaire*

The two methods of assignment, assignPOP and Locator, provided concordant results for two of the incursives (Fig 2). Inc-1 and Inc-2 had both clearly dispersed from A:Keriri, less than 10 km from where they were detected on Ngurapai (Fig 2: **×**). In assignPOP this was demonstrated by the high posterior (0.90 and 0.77) and relative (61.3 and 16.8) probabilities of assignment to A:Keriri. For Locator each incursive was placed among the A:Keriri genotypes for the clear

Author Manuscript

This article is protected by copyright. All rights reserved

majority of the 1000 bootstrap subsamples (Fig 2). Point estimates from Locator placed Inc-1 (10.5562° S, 142.2197° E) and Inc-2 (10.5585° S, 142.2173° E) amidst the A:Keriri samples.

Inc-3 and Inc-4 were assigned differently by assignPOP and Locator. This was not likely due to missing data (3.8% and 8.3%), as successful assignments with assignPOP have been recorded with > 40% missing data (Schmidt et al., 2019). assignPOP selected distant L:Mer (Inc-3) and I:Masig (Inc-4) as the most likely sources, though posterior (0.35 and 0.16) and relative (2.8 and 1.1) probabilities were low. Locator point estimates placed Inc-3 (10.2086° S, 142.2910° E) on Moa Island between the B:St Pauls and C:Kubin genotypes. Inc-4 (9.9216° S, 143.1598° E) was placed in the Torres Strait near I:Masig. Bootstrap subsampling revealed higher uncertainty for Inc-3 than for Inc-1 and Inc-2 but nevertheless indicated Moa Island was the likely source. No clear assignment was forthcoming for Inc-4.

International gene flow into the Torres Strait

Genome-wide gene flow

The *fineRADstructure* plot (Fig 3) shows coancestry between genotype pairs from TSI and other selected locations. See Figure S31 for the figure at full size, and Figure S32 for a PCA-UMAP ordination of the same genotypes using the same colour symbology. TSI genotypes were only weakly clustered by location, as seen by the poor 'sorting' of genotypes in the colour panel. Only one village (J:Ugar) had every genotype resolved into a single lineage. However, TSI genotypes formed a lineage with Indonesia and Timor-Leste (Fig 3: solid black square, top right) distinct from all other genotypes (solid black square, bottom left).

Within these broad groups were subgroups (Fig 3: dashed squares) placing East and Southeast Asian genotypes together with Fiji and Indonesian genotypes with Timor-Leste, as expected from previous analysis (Schmidt, Chung, Honnen, et al., 2020; Schmidt, Chung, van Rooyen, et

This article is protected by copyright. All rights reserved

Author Manuscript

al., 2020). PNG genotypes grouped with Vanuatu, suggesting recent coancestry. While the PNG and Vanuatu lineage had slightly greater coancestry with Southeast Asia than East Asia, this lineage was clearly differentiated from all others. PCA-UMAP confirmed the three major groups of TSI/Indonesia/Timor-Leste, East Asia/Southeast Asia/Fiji, and PNG/Vanuatu (Figure S32).

Gene flow from PNG into the TSI was suggested by the higher coancestry between PNG and some TSI genotypes, particularly those from L:Mer (Fig 3: blue asterisks). PNG also had higher coancestry with Jakarta and Bandung in Indonesia but not with Bali. While there was some evidence of coancestry between East Asian and TSI genotypes, this was also observed in East Asian and Indonesian genotypes, particularly those from Jakarta and Bandung. It was thus unclear whether recent gene flow had occurred from East Asia into the TSI or whether past gene flow from East Asia into Indonesia had introduced alleles that were then brought to the TSI during invasion.

Table 1 lists the number of private alleles in each village with and without the PNG samples included in filtering. When PNG was included, the greatest 'loss' of private alleles was observed in the eastern islands of I:Masig (30% loss), J:Ugar (35% loss), K:Erub (32% loss), and L:Mer (48% loss) and northern island of F:Iama (29% loss). These 'lost' alleles are identical by state and likely also identical by descent to those in PNG, indicating gene flow between each island and PNG. Notably, the fact that these alleles were not found on other islands suggests that this gene flow was also specific to each island, and thus is not due to one single incursion into the TSI followed by local spread.

Adaptive introgression

Outlier analysis revealed four genomic regions of interest, each containing irregularly structured SNPs spread across multiple RADtags within 1 Mbp, and detected as outliers by both *pcadapt* and sNMF. For pcadapt, a scree plot of the proportion of explained variance did not indicate an

This article is protected by copyright. All rights reserved

Author Manuscript

optimum K (Figure S33). However, visual inspection of scatterplots of the first eight PCs indicated that genetic structure among villages was much less apparent after the first four PCs (Figures S34-37). Accordingly, we restricted our analyses to $2 \leq K \leq 4$ for both programs.

The four genomic regions and all genes of known function contained therein are listed in Tables S3-S6. One of these regions, consisting of four SNPs across two RADtags ~330 kbp apart on scaffold NW_021838465.1, contained 12 genes of interest (Table S3). These included three cytochrome P450 9e2-like genes and nine probable cytochrome P450 9f2 genes. Cytochrome P450 9e2 genes have been linked to insecticide resistance and demonstrated strong upregulation in resistant *Ae. aegypti* mosquitoes (Kim Lien et al., 2019). Cytochrome P450 9f2 genes have also been linked to insecticide resistance (Etebari et al., 2018). The four SNPs enclosing the region were detected as outliers for all $2 \leq K \leq 4$ settings and were detected in both the imputed and unimputed datasets, which likely reflects the low missing data ($\bar{x} = 5.1\%$) at these SNPs.

We investigated these four SNPs more closely using the unimputed dataset from the entire Indo-Pacific region (Table 3). Non-reference alleles at these SNPs were found most frequently as a single linked haplotype, which in the TSI was detected at highest frequency on I:Masig (2018 and 2019) but also on A:Keriri and L:Mer. Outside of the TSI this haplotype was found only in Port Moresby in PNG. Considering non-reference alleles at the four SNPs individually, outside of the TSI and Port Moresby the only observation was a single individual from Singapore, which had two alleles identical by state at one of the SNPs (Table 3). Tables S7 and S8 list unimputed genotypes at these SNPs for all TSI individuals (Table S7) and all non-TSI individuals (Table S8). These show clear but imperfect patterns of linkage between the two RADtags, and clear geographical restrictions in where the alleles were found.

Author Manuscript

This article is protected by copyright. All rights reserved

Discussion

Spatial population genomics provides a powerful framework for investigating recent invasions. Our spatially consistent sampling within villages and simultaneous sampling across islands has permitted a range of analyses appropriate for characterising dispersal and adaptation in the recent *Ae. albopictus* invasion of the Torres Strait. Our findings point to a highly interconnected invasive metapopulation with frequent dispersal among both adjacent and non-adjacent islands, including incursions into uninvaded islands below the *cordon sanitaire*. The connectivity of this system also extends beyond borders, with evidence of gene flow from PNG into several of the nearest islands. Some of this gene flow was strongly suggestive of adaptive introgression, specifically at a set of linked outlier SNPs surrounding a genomic region containing 12 cytochrome P450 genes, of types (9e2 and 9f2) previously associated with insecticide resistance (Etebari et al., 2018; Kim Lien et al., 2019). Non-reference alleles at these outlier SNPs were found in four TSI villages, but not in any of the Indonesian populations from which the original invasion was sourced, or any population other than Port Moresby (PNG south coast), indicating a potential cross-country 'genetic invasion' of resistance alleles introgressed from southern PNG to the TSI. Broad-scale patterns of passive dispersal across islands contrasted with fine-scale patterns of active dispersal within villages, where spatial genetic structure was locally variable but displayed general trends of isolation by distance and neighbourhood size ($N_{W:}$ Wright, 1946) of 232–577.

When geographical genetic structure is weak, as in the TSI (Fig 1, Fig 3), this is often read as a sign of high gene flow between subregions (Bossart & Prowell, 1998). However, in new invasions this conclusion can be confounded by high regional coancestry stemming from the original invasion (Cristescu, 2015; Fitzpatrick et al., 2012). This study instead investigated dispersal by analysing close kin dyads, incursions below the *cordon sanitaire*, and gene flow from PNG. Kinship and incursion tracing both reveal dispersal patterns over the past ~1-2

This article is protected by copyright. All rights reserved

Author Manuscript

generations, and these methods should not be confounded by regional coancestry as they were either calibrated against estimates from within villages (kinship) or traced incursive mosquitoes to anywhere in continuous space within the region (incursions: Locator). Of the four incursives, only Inc-4 was likely confounded by coancestry, as seen by its central placement among the training genotypes (Fig 2). Likewise, international gene flow had no trouble with coancestry as the source of gene flow (PNG) was genetically very distinct from the source of the initial invasion (Indonesia) (Fig 3, Figure S32). Together these analyses suggest that long-distance dispersal is frequent throughout the region, though movement between nearby islands may be more common than between distant islands (Fig 2, Table 2). Despite evidence of direct gene flow from PNG (Table 1), the TSI retains a distinct genetic background (Fig 3), indicating that if any local eradication has occurred as postulated in Muzari et al. (2017) these islands would have been recolonised primarily from within the TSI. The consistent clustering of PNG and Vanuatu genotypes indicate the Vanuatuan *Ae. albopictus* invasion was likely established from PNG or another location of that lineage like the Solomon Islands (see Maynard et al., 2017).

These patterns of variation are similar to those found in some biological invasions but contrast with others where variation reflects ongoing founder events as a colonising species spreads. Similar patterns to those observed here have been established for western flower thrips where high rates of ongoing gene flow mediated by human transport have been hypothesised (Cao et al., 2017). In contrast, regional genetic patterns reflective of repeated colonisations with small sample sizes have been documented in other thrips where distant colonisation into vacant habitat is likely to have occurred (Cao et al., 2019), as well as where there have been intensive campaigns with chemical insecticides to subdue populations (Cao et al., 2017). For TSI *Ae. albopictus*, it appears that human-mediated migration has had a stronger effect on population structure than control efforts aiming to suppress populations. In this sense, the TSI *Ae. albopictus* appear to be structured as a metapopulation similar to classical metapopulations like Glanville fritillary butterflies (Fountain et al., 2018; Hanski et al., 2017).

Author Manuscript

This article is protected by copyright. All rights reserved

572 Although movement between islands has been suggested in previous studies (Beebe et al.,
573 2013; Maynard et al., 2017), this study provides the first evidence of dispersal between specific
574 islands. Our kinship analysis builds on previous uses of close kin to detect recent dispersal (M. Z.
575 Chen et al., 2020; Escoda et al., 2019, 2017; Fountain et al., 2018; Jasper et al., 2019; Schmidt et
576 al., 2018, 2017; Trense et al., 2020), and here demonstrates the use of regional genetic distance
577 patterns to also infer the direction of movement (Fig 1, Table 2). Inferring movement patterns
578 across single generations in this manner is becoming increasingly popular in applied work on
579 pests and conservation, where recent changes in demography and genetic structure mean that
580 assumptions of equilibrium demography are invalid and where anthropogenic impacts are
581 observed from one generation to the next. First order kin detected across islands were likely
582 transported via aeroplane, barge or small vessel movement. Although aircraft disinsection
583 (treatment with insecticides) is currently compulsory for all air travel between the TSI region
584 and the mainland of Australia, flights among the islands are not always disinsected. Future
585 research could investigate the relative contributions of different transport networks to *Ae.*
586 *albopictus* dispersal between islands and into the TSI from PNG.

587 The power of higher-resolution markers to better delineate geographical patterns among
588 genotypes is well-established (Escoda et al., 2017; Puckett & Eggert, 2016; Rašić et al., 2014).
589 While previous investigation of this system with lower-resolution markers detected some spatial
590 structuring among islands (Maynard et al., 2017), this study found no greater sorting of
591 genotypes by location despite higher marker power and despite sampling in this study having
592 taken place 3-11 years later (Fig 1, Fig 3). These results point to the TSI *Ae. albopictus*
593 population remaining highly interconnected over time. This interconnection is likely the reason
594 for high genetic variability within villages, including the inconsistent spatial structuring of
595 genetic variation among dyads (Table 1). Nevertheless, our estimate of N_W in villages (232 –
596 577) is consistent with a previous estimate for an urban *Ae. aegypti* population (268; Jasper et
597 al., 2019). Considering high densities of *Ae. albopictus* in the TSI, this estimate of N_W is
598 suggestive of some spatial structure within villages.

This article is protected by copyright. All rights reserved

Author Manuscript

Despite high regional gene flow, we managed to locate the source of 2-3 of the 4 incursions below the *cordon sanitaire*. Considering the absence of discrete populations within the TSI, the genotype-based method of Locator (Battey et al., 2020) was better suited to this dataset than the allele-frequency based method of assignPOP (K.-Y. Chen et al., 2018), though the latter still performed well for Inc-1 and Inc-2 and both methods successfully traced samples collected 10 months after the initial collections. Although Locator may best be suited to non-clustered sampling designs that are rare in island systems, Locator may sidestep issues relating to invasion age and frequent gene flow by not requiring that reference genotypes be sorted into populations. Locator's placement of Inc-4 across a wide swathe of the sampling range (Fig 2) indicated that this sample could not be assigned, which was not likely due to missing data but due to recent ancestry from multiple locations. If we treat the placement of Inc-4 as the expected outcome of a failure to assign, then the placement of Inc-3 on Moa Island with closely-clustered bootstrap replicates suggests that this is its likely origin. While population assignment methods such as assignPOP perform well when determining incursion sources pre-invasion (Schmidt, Chung, van Rooyen, et al., 2020; Schmidt et al., 2019), Locator appears a more useful tool for tracing local incursions after an initial invasion.

Detecting adaptive introgression is a powerful application of genomic datasets. For invasive systems adaptive introgression can be classed as a type of genetic invasion, in which alleles conferring a selective advantage that may interfere with control strategies are introduced after an initial invasion (Endersby-Harshman et al., 2020; Riveron et al., 2013; Schmidt et al., 2019). Our detection of a set of alleles from PNG that have introgressed into the TSI is strongly suggestive of a genetic invasion. In the TSI, introgressed alleles at the four outlier SNPs were found on four of the islands, but were particularly common on I:Masig, where their frequency remained stable between 2018 and 2019. Considering the current low allele frequencies on most of the islands, the imperfect linkage between the alleles, and the abandonment of the insecticide program in the TSI a decade previously (Muzari et al., 2017), our observations may reflect the aftermath of a genetic invasion that spread into the TSI during the height of the insecticide program. During this time, I:Masig may have been acting as an 'invasive bridgehead'

This article is protected by copyright. All rights reserved

Author Manuscript

(Estoup & Guillemaud, 2010) for the genetic invasion, from which resistance alleles spread to other islands.

While traditionally 'introgression' refers to the interspecific transfer of genetic material, which has also been observed in mosquitoes (Norris et al., 2015), introgression between differentiated populations as observed here may be more common and may allow for faster spread of advantageous alleles if interspecific mating is rare or confers greater fitness costs. Considering also the connectivity of the TSI/PNG system, where dispersal takes place among distant islands and across countries, if advantageous alleles arise in any given location they may rapidly spread throughout the region if similar selective pressures are present. This rapid spread of alleles may be a threat in any biological system with high dispersal rates and common selective pressures. In the TSI, linkage was strong between the two RADtags surrounding the 12 cytochrome P450 genes, though not as strong as the near-perfect linkage observed around pyrethroid resistance mutations in *Ae. aegypti* (Endersby-Harshman et al., 2020). This may reflect relatively weaker selection around the cytochrome P450 genes given that area-wide insecticide use in the TSI was phased out ~10 years before sampling for this project (Muzari et al., 2017) and that in PNG most documented insecticide use occurs outside of cities like Port Moresby (Demok et al., 2019). While recent bioassays on PNG *Ae. albopictus* demonstrated resistance to DDT, these investigated only Madang *Ae. albopictus* (Demok et al., 2019), though emerging results indicate these resistance patterns are similar across PNG (unpublished data). Nevertheless, the detected introgression was evidently not from Madang but from either Port Moresby or an unsampled location in PNG. Future work is needed to determine what role the 12 cytochrome P450 genes have in conferring insecticide resistance in this system and in *Ae. albopictus* more broadly.

Conclusions

This article is protected by copyright. All rights reserved

Author Manuscript

Dispersal can be difficult to investigate in many biological systems, particularly those subjected to different ongoing dispersal processes. The recent biological invasion of *Ae. albopictus* into TSI provides an example of such a system, and here we have performed a comprehensive characterisation of dispersal using temporally-restricted sampling and an appropriate set of population genomic analyses. We detected recent dispersal between distant and adjacent islands and found no increase in spatial genetic structure over time, suggesting that the invaded islands may best be considered a metapopulation connected by frequent gene flow, rather than individual 'populations'. We also made use of the strongly-differentiated PNG *Ae. albopictus* population to detect gene flow from PNG into specific islands, and found evidence that a genetic invasion of insecticide resistance alleles may have spread into the TSI from the south coast of PNG. These methods will be broadly applicable to other taxa, particularly those affected by anthropogenic processes. For *Ae. albopictus*, the findings inform strategies for the control of this globally-invasive pest across a broad range of spatial scales.

Acknowledgements

For sample collection we thank Esther Anderson, Moses Laman, Rotarians against Malaria PNG staff, and the PNGIMR Vector Borne Disease Unit entomology team. We thank Moshe Jasper for helpful discussions. For funding we thank Queensland Health, the Commonwealth Department of Health, the Far North Queensland Hospital Foundation (FNQHF), and the NHMRC (Programme Grant 1037003, Career Development Fellowship 1141441, and Fellowship Grants 1118640 and 5121190). T.S. was financially supported by a Australian Government Research Training Program Scholarship.

Data Accessibility

This article is protected by copyright. All rights reserved

Author Manuscript

Aligned .bam sequence files for 331 TSI *Ae. albopictus*, 4 incursive *Ae. albopictus*, and 36 PNG *Ae. albopictus* are available through the Sequence Read Archive at NCBI Genbank: PRJNA684450.

Author Contributions

TLS, TS, GE, SR, and AAH designed the study. TS, SK, SD, QY, MOM, GE, MB, RB and PH collected the data. TLS, TS, JC and MAF analysed the data. TLS, MAF, GE, TRB, SR and AAH provided supervision. TLS, TS and AAH wrote the paper with edits from all authors.

References

Battey, C. J., Ralph, P. L., & Kern, A. D. (2020). Predicting geographic location from genetic variation with deep neural networks. *ELife*, *9*, 1–22. doi: 10.7554/eLife.54507

Beebe, N. W., Ambrose, L., Hill, L. A., Davis, J. B., Hapgood, G., Cooper, R. D., … van den Hurk, A. F. (2013). Tracing the Tiger: Population Genetics Provides Valuable Insights into the *Aedes* (*Stegomyia*) *albopictus* Invasion of the Australasian Region. *PLoS Neglected Tropical Diseases*, *7*(8), e2361. doi: 10.1371/journal.pntd.0002361

Bossart, J. L., & Prowell, D. P. (1998). Genetic estimates of population structure and gene flow: Limitations, lessons and new directions. *Trends in Ecology and Evolution*, *13*(5), 202–206. doi: 10.1016/S0169-5347(97)01284-6

Bradburd, G. S., & Ralph, P. L. (2019). Spatial Population Genetics: It's about Time. *Annual Review of Ecology, Evolution, and Systematics*, *50*(1), 427–449. doi: 10.1146/annurev-ecolsys-110316-022659

Browning, B. L., & Browning, S. R. (2016). Genotype Imputation with Millions of Reference Samples. *The American Journal of Human Genetics*, *98*(1), 116–126. doi: 10.1016/J.AJHG.2015.11.020

Cao, L., Gao, Y., Gong, Y., Chen, J., Chen, M., Hoffmann, A., & Wei, S. (2019). Population analysis reveals

This article is protected by copyright. All rights reserved

Author Manuscript

genetic structure of an invasive agricultural thrips pest related to invasion of greenhouses and suitable climatic space. *Evolutionary Applications*, *12*(10), 1868–1880. doi: 10.1111/eva.12847

Cao, L. J., Wang, Z. H., Gong, Y. J., Zhu, L., Hoffmann, A. A., & Wei, S. J. (2017). Low genetic diversity but strong population structure reflects multiple introductions of western flower thrips (Thysanoptera: Thripidae) into China followed by human-mediated spread. *Evolutionary Applications*, *10*(4), 391–401. doi: 10.1111/eva.12461

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. doi: 10.1111/mec.12354

Chen, K.-Y., Marschall, E. A., Sovic, M. G., Fries, A. C., Gibbs, H. L., & Ludsin, S. A. (2018). *assignPOP*: An R package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. *Methods in Ecology and Evolution*, *9*(2), 439–446. doi: 10.1111/2041-210X.12897

Chen, M. Z., Cao, L. J., Li, B. Y., Chen, J. C., Gong, Y. J., Yang, Q., … Wei, S. J. (2020). Migration trajectories of the diamondback moth *Plutella xylostella* in China inferred from population genomic variation. *Pest Management Science*, ps.6188. doi: 10.1002/ps.6188

Combs, M., Puckett, E. E., Richardson, J., Mims, D., & Munshi-South, J. (2018). Spatial population genomics of the brown rat (*Rattus norvegicus*) in New York City. *Molecular Ecology*, *27*(1), 83–98. doi: 10.1111/mec.14437

Conomos, M. P., Reiner, A. P., Weir, B. S., & Thornton, T. A. (2016). Model-free Estimation of Recent Genetic Relatedness. *American Journal of Human Genetics*, *98*(1), 127–148. doi: 10.1016/j.ajhg.2015.11.022

Cristescu, M. E. (2015). Genetic reconstructions of invasion history. *Molecular Ecology*, *24*(9), 2212–2225. doi: 10.1111/mec.13117

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. doi: 10.1093/bioinformatics/btr330

Demok, S., Endersby-Harshman, N., Vinit, R., Timinao, L., Robinson, L. J., Susapu, M., … Karl, S. (2019).

This article is protected by copyright. All rights reserved

Author Manuscript

727 Insecticide resistance status of *Aedes aegypti* and *Aedes albopictus* mosquitoes in Papua New
728 Guinea. *Parasites and Vectors*, *12*(1), 333. doi: 10.1186/s13071-019-3585-6

729 Diaz-Papkovich, A., Anderson-Trocmé, L., Ben-Eghan, C., & Gravel, S. (2019). UMAP reveals cryptic
730 population structure and phenotype heterogeneity in large genomic cohorts. *PLoS Genetics*, *15*(11).
731 doi: 10.1371/journal.pgen.1008432

732 Dray, S., & Dufour, A. B. (2007). The ade4 package: Implementing the duality diagram for ecologists.
733 *Journal of Statistical Software*, *22*(4), 1–20. doi: 10.18637/jss.v022.i04

734 Endersby-Harshman, N. M., Schmidt, T. L., Chung, J., van Rooyen, A., Weeks, A. R., & Hoffmann, A. A.
735 (2020). Heterogeneous genetic invasions of three insecticide resistance mutations in Indo-Pacific
736 populations of *Aedes aegypti* (L.). *Molecular Ecology*, *29*(9), 1628–1641. doi: 10.1111/mec.15430

737 Escoda, L., Fernández-González, Á., & Castresana, J. (2019). Quantitative analysis of connectivity in
738 populations of a semi-aquatic mammal using kinship categories and network assortativity.
739 *Molecular Ecology Resources*, *19*(2), 310–326. doi: 10.1111/1755-0998.12967

740 Escoda, L., González-Esteban, J., Gómez, A., & Castresana, J. (2017). Using relatedness networks to infer
741 contemporary dispersal: Application to the endangered mammal *Galemys pyrenaicus*. *Molecular*
742 *Ecology*, *26*(13), 3343–3357. doi: 10.1111/mec.14133

743 Estoup, A., & Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: Why, how and
744 so what? *Molecular Ecology*, *19*(19), 4113–4130. doi: 10.1111/j.1365-294X.2010.04773.x

745 Etebari, K., Afrad, M. H., Tang, B., Silva, R., Furlong, M. J., & Asgari, S. (2018). Involvement of microRNA
746 miR-2b-3p in regulation of metabolic resistance to insecticides in *Plutella xylostella*. *Insect*
747 *Molecular Biology*, *27*(4), 478–491. doi: 10.1111/imb.12387

748 Fitzpatrick, B. M., Fordyce, J. A., Niemiller, M. L., & Reynolds, R. G. (2012). What can DNA tell us about
749 biological invasions? *Biological Invasions*, *14*(2), 245–253. doi: 10.1007/s10530-011-0064-1

750 Fitzpatrick, B. M., Johnson, J. R., Kump, D. K., Smith, J. J., Voss, S. R., & Shaffer, H. B. (2010). Rapid spread
751 of invasive genes into a threatened native species. *Proceedings of the National Academy of Sciences*
752 *of the United States of America*, *107*(8), 3606–3610. doi: 10.1073/pnas.0911802107

753 Fountain, T., Husby, A., Nonaka, E., DiLeo, M. F., Korhonen, J. H., Rastas, P., … Hanski, I. (2018). Inferring

This article is protected by copyright. All rights reserved

Author Manuscript

754 dispersal across a fragmented landscape using reconstructed families in the Glanville fritillary
755 butterfly. *Evolutionary Applications*, *11*(3), 287–297. doi: 10.1111/eva.12552

756 Frichot, E., & François, O. (2015). LEA : An R package for landscape and ecological association studies.
757 *Methods in Ecology and Evolution*, *6*(8), 925–929. doi: 10.1111/2041-210X.12382

758 Gattepaille, L. M., Jakobsson, M., & Blum, M. G. B. (2013, May 20). Inferring population size changes with
759 sequence and SNP data: Lessons from human bottlenecks. *Heredity*, Vol. 110, pp. 409–419. Nature
760 Publishing Group. doi: 10.1038/hdy.2012.120

761 Gillespie, R. G., Baldwin, B. G., Waters, J. M., Fraser, C. I., Nikula, R., & Roderick, G. K. (2012). Long-
762 distance dispersal: a framework for hypothesis testing. *Trends in Ecology and Evolution*, *27*(1), 47–
763 56. doi: 10.1016/j.tree.2011.08.009

764 Goubert, C., Henri, H., Minard, G., Valiente Moro, C., Mavingui, P., Vieira, C., & Boulesteix, M. (2017).
765 High-throughput sequencing of transposable element insertions suggests adaptive evolution of the
766 invasive Asian tiger mosquito towards temperate environments. *Molecular Ecology*, *26*(15), 3968–
767 3981. doi: 10.1111/mec.14184

768 Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular*
769 *Ecology Notes*, *5*(1), 184–186. doi: 10.1111/j.1471-8286.2004.00828.x

770 Hanski, I., Schulz, T., Wong, S. C., Ahola, V., Ruokolainen, A., & Ojanen, S. P. (2017). Ecological and
771 genetic basis of metapopulation persistence of the Glanville fritillary butterfly in fragmented
772 landscapes. *Nature Communications*, *8*(1), 1–11. doi: 10.1038/ncomms14504

773 Hardy, O. J., & Vekemans, X. (2002). SPAGeDi: a versatile computer program to analyse spatial genetic
774 structure at the individual or population levels. *Molecular Ecology Notes*, *2*(4), 618–620. doi:
775 10.1046/j.1471-8286.2002.00305.x

776 Harrington, L. C., Scott, T. W., Lerdthusnee, K., Coleman, R. C., Costero, A., Clark, G. G., … Edman, J. D.
777 (2005). Dispersal of the dengue vector *Aedes aegypti* within and between rural communities. *The*
778 *American Journal of Tropical Medicine and Hygiene*, *72*(2), 209–220. doi:
779 10.4269/ajtmh.2005.72.209

780 Hengeveld, R. (1989). *Dynamics of biological invasions*. London: Chapman & Hall.

Author Manuscript

This article is protected by copyright. All rights reserved

781 Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., … Whitlock, M. C.
782 (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future
783 directions. *American Naturalist*, *188*(4), 379–397. doi: 10.1086/688018

784 Howard, W. E. (1960). Innate and Environmental Dispersal of Individual Vertebrates. *American Midland*
785 *Naturalist*, *63*(1), 152–161. doi: 10.2307/2422936

786 Jasper, M., Schmidt, T. L., Ahmad, N. W., Sinkins, S. P., & Hoffmann, A. A. (2019). A genomic approach to
787 inferring kinship reveals limited intergenerational dispersal in the yellow fever mosquito. *Molecular*
788 *Ecology Resources*, *19*(5), 1254–1264. doi: 10.1111/1755-0998.13043

789 Johansen, C. A., Van Den Hurk, A. F., Ritchie, S. A., Zborowski, P., Nisbet, D. J., Paru, R., … Mackenzie, J. S.
790 (2000). Isolation of Japanese encephalitis virus from mosquitoes (Diptera: Culicidae) collected in
791 the western province of Papua New Guinea, 1997-1998. *American Journal of Tropical Medicine and*
792 *Hygiene*, *62*(5), 631–638. doi: 10.4269/ajtmh.2000.62.631

793 Jost, L. (2008). GST and its relatives do not measure differentiation. *Molecular Ecology*, *17*(18), 4015–
794 4026. doi: 10.1111/j.1365-294X.2008.03887.x

795 Kim Lien, N. T., Hong Ngoc, N. T., Lan, N. N., Hien, N. T., Van Tung, N., Thanh Ngan, N. T., … Huong Binh,
796 N. T. (2019). Transcriptome Sequencing and Analysis of Changes Associated with Insecticide
797 Resistance in the Dengue Mosquito (*Aedes aegypti*) in Vietnam. *American Journal of Tropical*
798 *Medicine and Hygiene*, *100*(5), 1240–1248. doi: 10.4269/ajtmh.18-0607

799 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4),
800 357–359. doi: 10.1038/nmeth.1923

801 Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure
802 inference with genomic data sets. *Molecular Ecology Resources*, *19*(3), 639–647. doi: 10.1111/1755-
803 0998.12995

804 Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt : an R package to perform genome scans for selection
805 based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77. doi:
806 10.1111/1755-0998.12592

807 Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure:

This article is protected by copyright. All rights reserved

Author Manuscript

808 Population Inference from RADseq Data. *Molecular Biology and Evolution*, *35*(5), 1284–1290. doi:
809 10.1093/molbev/msy023

810 Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust
811 relationship inference in genome-wide association studies. *Bioinformatics*, *26*(22), 2867–2873. doi:
812 10.1093/bioinformatics/btq559

813 Maynard, A. J., Ambrose, L., Cooper, R. D., Chow, W. K., Davis, J. B., Muzari, M. O., … Beebe, N. W.
814 (2017). Tiger on the prowl: Invasion history and spatio-temporal genetic structure of the Asian tiger
815 mosquito *Aedes albopictus* (Skuse 1894) in the Indo-Pacific. *PLOS Neglected Tropical Diseases*,
816 *11*(4), e0005546. doi: 10.1371/journal.pntd.0005546

817 Mcfarlane, J. (1998). Torres Strait: Policing the Open Border. *Trends and Issues in Crime and Criminal*
818 *Justice*, 1–6. Canberra: Australian Institute of Criminology. Retrieved from http://www.aic.gov.au

819 Muzari, M. O., Devine, G., Davis, J., Crunkhorn, B., van den Hurk, A., Whelan, P., … Ritchie, S. (2017).
820 Holding back the tiger: Successful control program protects Australia from *Aedes albopictus*
821 expansion. *PLOS Neglected Tropical Diseases*, *11*(2), e0005286. doi: 10.1371/journal.pntd.0005286

822 Norris, L. C., Main, B. J., Lee, Y., Collier, T. C., Fofana, A., Cornel, A. J., & Lanzaro, G. C. (2015). Adaptive
823 introgression in an African malaria mosquito coincident with the increased usage of insecticide-
824 treated bed nets. *Proceedings of the National Academy of Sciences of the United States of America*,
825 *112*(3), 815–820. doi: 10.1073/pnas.1418892112

826 Palatini, U., Masri, R. A., Cosme, L. V., Koren, S., Thibaud-Nissen, F., Biedler, J. K., … Bonizzoni, M. (2020).
827 Improved reference genome of the arboviral vector *Aedes albopictus*. *BioRxiv*, 2020.02.28.969527.
828 doi: 10.1101/2020.02.28.969527

829 Pélissié, B., Crossley, M. S., Cohen, Z. P., & Schoville, S. D. (2018, April 1). Rapid evolution in insect pests:
830 the importance of space and time in population genomics studies. *Current Opinion in Insect Science*,
831 Vol. 26, pp. 8–16. Elsevier Inc. doi: 10.1016/j.cois.2017.12.008

832 Puckett, E. E., & Eggert, L. S. (2016). Comparison of SNP and microsatellite genotyping panels for spatial
833 assignment of individuals to natal range: A case study using the American black bear (*Ursus*
834 *americanus*). *Biological Conservation*, *193*, 86–93. doi: 10.1016/J.BIOCON.2015.11.020

Author Manuscript

This article is protected by copyright. All rights reserved

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, *16*(3), 608–627. doi: 10.1111/1755-0998.12512

Rašić, G., Filipović, I., Weeks, A. R., & Hoffmann, A. A. (2014). Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, *Aedes aegypti*. *BMC Genomics*, *15*(1), 275. doi: 10.1186/1471-2164-15-275

Ritchie, S. A., Moore, P., Carruthers, M., Williams, C., Montgomery, B., Foley, P., … Russell, R. C. (2006). Discovery of a Widespread Infestation of *Aedes albopictus* in the Torres Strait, Australia. *Journal of the American Mosquito Control Association*, *22*(3), 358–365. doi: 10.2987/8756-971X(2006)22[358:DOAWIO]2.0.CO;2

Riveron, J. M., Irving, H., Ndula, M., Barnes, K. G., Ibrahim, S. S., Paine, M. J. I., & Wondji, C. S. (2013). Directionally selected cytochrome P450 alleles are driving the spread of pyrethroid resistance in the major malaria vector *Anopheles funestus*. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(1), 252–257. doi: 10.1073/pnas.1216705110

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011, January 1). Integrative genomics viewer. *Nature Biotechnology*, Vol. 29, pp. 24–26. Nature Publishing Group. doi: 10.1038/nbt.1754

Rousset, F. (2000). Genetic differentiation between individuals. *Journal of Evolutionary Biology*, *13*(1), 58–62. doi: 10.1046/j.1420-9101.2000.00137.x

Schmidt, T. L., Chung, J., Honnen, A. C., Weeks, A. R., & Hoffmann, A. A. (2020). Population genomics of two invasive mosquitoes (*Aedes aegypti* and *Aedes albopictus*) from the Indo-Pacific. *PLoS Neglected Tropical Diseases*, *14*(7), 1–24. doi: 10.1371/journal.pntd.0008463

Schmidt, T. L., Chung, J., van Rooyen, A. R., Sly, A., Weeks, A. R., & Hoffmann, A. A. (2020). Incursion pathways of the Asian tiger mosquito (Aedes albopictus) into Australia contrast sharply with those of the yellow fever mosquito (Aedes aegypti). *Pest Management Science*, *76*(12), 4202–4209. doi: 10.1002/ps.5977

Schmidt, T. L., Filipović, I., Hoffmann, A. A., & Rašić, G. (2018). Fine-scale landscape genomics helps explain the slow spatial spread of *Wolbachia* through the *Aedes aegypti* population in Cairns,

This article is protected by copyright. All rights reserved

Author Manuscript

Australia. *Heredity*, *120*(5), 386–395. doi: 10.1038/s41437-017-0039-9

Schmidt, T. L., Rašić, G., Zhang, D., Zheng, X., Xi, Z., & Hoffmann, A. A. (2017). Genome-wide SNPs reveal the drivers of gene flow in an urban population of the Asian Tiger Mosquito, *Aedes albopictus*. *PLOS Neglected Tropical Diseases*, *11*(10), e0006009. doi: 10.1371/journal.pntd.0006009

Schmidt, T. L., van Rooyen, A. R., Chung, J., Endersby-Harshman, N. M., Griffin, P. C., Sly, A., … Weeks, A. R. (2019). Tracking genetic invasions: Genome-wide single nucleotide polymorphisms reveal the source of pyrethroid-resistant *Aedes aegypti* (yellow fever mosquito) incursions at international ports. *Evolutionary Applications*, *12*(6), 1136–1146. doi: 10.1111/eva.12787

Schultz, C. B., & Crone, E. E. (2001). Edge-mediated dispersal behavior in a prairie butterfly. *Ecology*, *82*(7), 1879–1892. doi: 10.1890/0012-9658(2001)082[1879:EMDBIA]2.0.CO;2

Sharov, A. A., & Liebhold, A. M. (1998). Model of slowing the spread of gypsy moth (Lepidoptera: Lymantriidae) with a barrier zone. *Ecological Applications*, *8*(4), 1170–1179. doi: 10.1890/1051-0761(1998)008[1170:MOSTSO]2.0.CO;2

Sherpa, S., Blum, M. G. B., Capblancq, T., Cumer, T., Rioux, D., & Després, L. (2019). Unravelling the invasion history of the Asian tiger mosquito in Europe. *Molecular Ecology*, *28*(9), 2360–2377. doi: 10.1111/mec.15071

Shi, S., Yuan, N., Yang, M., Du, Z., Wang, J., Sheng, X., … Xiao, J. (2018). Comprehensive Assessment of Genotype Imputation Performance. *Human Heredity*, *83*(3), 107–116. doi: 10.1159/000489758

Trense, D., Schmidt, T. L., Yang, Q., Chung, J., Hoffmann, A. A., & Fischer, K. (2020). Anthropogenic and natural barriers affect genetic connectivity in an Alpine butterfly. *Molecular Ecology*. doi: 10.1111/mec.15707

van den Hurk, A. F., Nicholson, J., Beebe, N. W., Davis, J., Muzari, O. M., Russell, R. C., … Ritchie, S. A. (2016). Ten years of the Tiger: *Aedes albopictus* presence in Australia since its discovery in the Torres Strait in 2005. *One Health*, *2*, 19–24. doi: 10.1016/J.ONEHLT.2016.02.001

Waters, J. M., Fraser, C. I., & Hewitt, G. M. (2013). Founder takes all: density-dependent processes structure biodiversity. *Trends in Ecology & Evolution*, *28*, 78–85. doi: 10.1016/j.tree.2012.08.024

Wright, S. (1943). Isolation by Distance. *Genetics*, *28*(2), 114–138. doi: 10.1016/B978-0-12-374984-

This article is protected by copyright. All rights reserved

Author Manuscript

0.00820-2

Wright, S. (1946). Isolation by distance under diverse systems of mating. *Genetics*, *31*(1), 39–59. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1209315/pdf/39.pdf

Yadav, S., Stow, A., & Dudaniec, R. Y. (2020). Microgeographic adaptation corresponds with elevational distributions of congeneric montane grasshoppers. *Molecular Ecology*, mec.15739. doi: 10.1111/mec.15739

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. doi: 10.1093/bioinformatics/bts606

Figure Captions

Fig 1: Locations and genetic structure of the TSI villages. A:Keriri, B:St Pauls (Moa Island) C:Kubin (Moa Island), D:Badu, E:Mabuiag, F:Iama, G:Warraber, H:Poruma, I:Masig, J:Ugar, K:Erub, L:Mer. Plots are of sparse non-negative matrix factorisation (sNMF) on the Region22 dataset (see Datasets and filtering), and are centred over each village. Grey and blue sections of each vertical bar represent ancestral lineages of individuals, assuming K = 2. White-dotted circles and black-dotted squares of the same colour indicate close kin dyads found across villages, with squares denoting the origin (see Table 2). The white crossed square indicates Dauan. The "×" indicates Ngurapai, where the four incursive mosquitoes were detected. Map inset shows the Australian mainland, Papua New Guinea (PNG) and the Torres Strait Islands (TSI). Maps produced in QGIS (v3.6) using the Torres Strait Clear sky Landsat (https://eatlas.org.au/data/uuid/71c8380e-4cdc-4544-98b6-8a5c328930ad) and the Bright Earth eAtlas basemap v1.0 (https://eatlas.org.au/data/uuid/ac57aa5a-233b-4c2c-bd52-1fb40a31). See Figures S1-S13 for village maps and sampling point locations, and Figures S15-S17 for sNMF results for $2 \leq K \leq 4$.

This article is protected by copyright. All rights reserved

Author Manuscript

Fig 2: Predicted locations of the four incursives from 1000 bootstrapping runs of Locator. Each prediction is depicted with a red circle. The "×" indicates Ngurapai, where the incursives were detected. The dotted line describes the position of the *cordon sanitaire*. See main text for point estimates of locations.

Fig 3: FineRADstructure coancestry map and tree. The left hand side panel indicates genotype sampling locations coloured by geographical proximity and with all non-TSI genotypes labelled. The central panel shows coancestry between genotypes, with light yellow indicating low coancestry, and darker yellows, oranges and reds indicating progressively higher coancestry. Solid black squares indicate the two major clades of TSI/Indonesia/Timor-Leste (top right) and all other genotypes (bottom left). Dashed squares indicate subgroups of PNG/Vanuatu, East Asia/Southeast Asia/Fiji, and Indonesia/Timor-Leste. The dotted square indicates J:Ugar, the only TSI village resolved into a single lineage. Blue asterisks indicate higher coancestry between PNG and TSI genotypes, particularly those from L:Mer. See Figure S31 for full-size figure.

Author Manuscript

This article is protected by copyright. All rights reserved

940

941

942 **Tables**

943

944 Table 1: TSI village details and population genetic parameter estimates including
945 heterozygosity (H_O) and nucleotide diversity (π) at variant sites, Tajima's D, Rousset's *a* for dyads
946 < 50 m and > 500 m apart, and number of private alleles with and without inclusion of PNG
947 samples.

Village key (Fig 1)	Village	Island area (km²)	Human pop size (2016)	H_O (±SE)	π (±SE)	Tajima's D (±SE)	Mean (±SD) Rousset's *a*: <50m	Mean (±SD) Rousset's *a*: >500m	Private alleles (no PNG)	Private alleles (PNG)
A	Keriri	14.5	268	0.063 (±0.001)	0.074 (±0.001)	-1.164 (±0.026)	0.25 (±0.07)	0.20 (±0.10)	118	98
B	St Pauls (Moa)	152.4	248	0.063 (±0.001)	0.074 (±0.001)	-1.010 (±0.030)	0.18 (±0.08)	0.20 (±0.10)	123	91
C	Kubin (Moa)	152.4	198	0.064 (±0.001)	0.075 (±0.001)	-0.809 (±0.032)	0.16 (±0.09)	0.16 (±0.08)	97	80
D	Badu	101.9	813	0.065 (±0.001)	0.078 (±0.001)	-1.164 (±0.028)	0.18 (±0.07)	0.21 (±0.09)	56	42
E	Mabuiag	6.3	210	0.065 (±0.001)	0.076 (±0.001)	-0.986 (±0.028)	0.15 (±0.11)	0.19 (±0.06)	146	115
F	Iama	1.5	319	0.063 (±0.001)	0.074 (±0.001)	-0.691 (±0.034)	0.21 (±0.08)	0.12 (±0.07)	120	85
G	Warraber	0.6	245	0.066 (±0.001)	0.078 (±0.001)	-0.980 (±0.027)	0.20 (±0.06)	0.23 (±0.05)	77	66
H	Poruma	0.4	167	0.065 (±0.001)	0.076 (±0.001)	-0.972 (±0.028)	0.17 (±0.08)	0.19 (±0.07)	105	83

Author Manuscript

This article is protected by copyright. All rights reserved

I	Masig (2018)	1.5	270	0.069 (±0.001)	0.080 (±0.001)	-1.054 (±0.027)	0.11 (±0.00)	0.19 (±0.05)	128	90
I	Masig (2019)			0.073 (±0.001)	0.082 (±0.001)	-0.908 (±0.029)	0.14 (±0.02)	0.14 (±0.04)		
J	Ugar	0.4	85	0.063 (±0.001)	0.074 (±0.001)	-1.106 (±0.029)	0.22 (±0.06)	0.19 (±0.05)	130	85
K	Erub	6.9	328	0.063 (±0.001)	0.076 (±0.001)	-1.209 (±0.027)	0.19 (±0.09)	0.24 (±0.08)	182	123
L	Mer	4.3	450	0.067 (±0.001)	0.081 (±0.001)	-1.240 (±0.026)	0.22 (±0.10)	0.25 (±0.09)	148	77

Table 2: Putative close kin dyads within and across villages. Within villages, the dyad at E:Mabuiag had the lowest PC-Relate kinship coefficient of the three close kin dyads, so this was used as the cut-off for designating close kin across villages. Mean kinship describes the mean of kinship coefficients between Sib-1 and Sib-2 and all other individuals from that village, using kinship coefficients from the 3 PC treatment. Higher mean kinship in Village-2 shows this is the origin.

Sib-1	Sib-2	Origin	Distance (km)	Kinship (KING)	Kinship (PC-Relate 3 PCs)	Kinship (PC-Relate 10 PCs)	Mean kinship Village-1 (±SD)	Mean kinship Village-2 (±SD)
Within villages								
K:Erub	K:Erub		0.17	0.43	0.59	0.47		
B:St Pauls	B:St Pauls		0	0.42	0.51	0.48		
E:Mabuiag	E:Mabuiag		0	0.34	0.32	0.28		
Between villages								
C:Kubin	A:Keriri	A:Keriri	35.26		0.52	0.35	-0.015 (±0.017)	0.042 (±0.038)
H:Poruma	G:Warraber	G:Warraber	31.51		0.51	0.41	-0.017 (±0.017)	0.032 (±0.033)

Author Manuscript

This article is protected by copyright. All rights reserved

C:Kubin	B:St Pauls	B:St Pauls	13.50	0.49	0.46	-0.036 (±0.018)	0.001 (±0.021)
L:Mer	E:Mabuiag	E:Mabuiag	202.63	0.44	0.42	-0.013 (±0.018)	0.055 (±0.052)
B:St Pauls	E:Mabuiag	E:Mabuiag	30.59	0.44	0.44	-0.002 (±0.016)	0.060 (±0.040)

954
955
956
957
958
959

960 Table 3: Frequencies of non-reference alleles (p) at the four outlier SNPs on scaffold
961 NW_021838465.1. RADtag1 had one outlier SNP (pos: 84554160), RADtag2 had three outlier
962 SNPs (pos: 84888984, 84888990, 84888991) which always cooccurred except in case of missing
963 data. See Table S3 for a list of genes of known function within this region and Tables S7-S8 for
964 genotypes of all individuals at the four outlier SNPs.
965

Location	**p(84554160) RADtag1**	**p(84888984) RADtag2**	**p(84888990) RADtag2**	**p(84888991) RADtag2**
TSI				
A:Keriri	0.056	0.045	0.045	0.045
I:Masig (2018)	0.238	0.272	0.272	0.272
I:Masig (2019)	0.250	0.250	0.250	0.250
L:Mer	0.040	0.021	0.021	0
J:Ugar	0	0.042	0.042	0.042
All other locations	0	0	0	0
International				

Author Manuscript

This article is protected by copyright. All rights reserved

Port Moresby	0.235	0.265	0.282	0.265
Singapore	0.071	0	0	0
All other locations	0	0	0	0

966

967

968 ## Supporting Information Legends

969

970 **Text**

971 Text S1: Details of the genotyping pipeline.

972

973

974 **Tables**

975 Table S1: Initial assignPOP run showing assignment probability to each TSI and non-TSI location,
976 and aggregated assignment probabilities for all TSI and all non-TSI locations.

977 Table S2: Pairwise FST estimates among TSI villages.

978 Table S3: The first of four genomic regions containing multiple outlier SNPs found on multiple
979 RADtags, and the genes of known product or function located between these RADtags.

980 Table S4: The second of four genomic regions containing multiple outlier SNPs found on
981 multiple RADtags, and the genes of known product or function located between these RADtags.

982 Table S5: The third of four genomic regions containing multiple outlier SNPs found on multiple
983 RADtags, and the genes of known product or function located between these RADtags.

984 Table S6: The fourth of four genomic regions containing multiple outlier SNPs found on multiple
985 RADtags, and the genes of known product or function located between these RADtags.

986 Table S7: Non-reference alleles at the four outlier SNPs in Region 1 (see Table S2) found in TSI
987 villages.

Author Manuscript

This article is protected by copyright. All rights reserved

Table S8: Non-reference alleles at the four outlier SNPs in Region 1 (see Table S2) found in non-TSI locations.

Figures

Figure S1: Locations of *Aedes albopictus* sweep net collections on Badu Island. Map produced using Mapinfo (2019) with the Queensland basemap satellite imagery (2020).

Figure S2: Locations of *Aedes albopictus* sweep net collections on Dauan Island. Map produced using Mapinfo (2019) with the Queensland basemap satellite imagery (2020).

Figure S3: Locations of *Aedes albopictus* sweep net collections on Erub Island. Map produced using Mapinfo (2019) with the Queensland basemap satellite imagery (2020).

Figure S4: Locations of *Aedes albopictus* sweep net collections on Iama Island. Map produced using Mapinfo (2019) with the Queensland basemap satellite imagery (2020).

Figure S5: Locations of *Aedes albopictus* sweep net collections on Keriri Island. Map produced using Mapinfo (2019) with the Queensland basemap satellite imagery (2020).

Figure S6: Locations of *Aedes albopictus* sweep net collections on Mabuiag Island. Map produced using Mapinfo (2019) with the Queensland basemap satellite imagery (2020).

Figure S7: Locations of *Aedes albopictus* sweep net collections on Masig Island. Map produced using Mapinfo (2019) with the Queensland basemap satellite imagery (2020).

Figure S8: Locations of *Aedes albopictus* sweep net collections on Poruma Island. Map produced using Mapinfo (2019) with the Queensland basemap satellite imagery (2020).

Figure S9: Locations of *Aedes albopictus* sweep net collections on Mer Island. Map produced using Mapinfo (2019) with the Queensland basemap satellite imagery (2020).

Figure S10: Locations of *Aedes albopictus* sweep net collections on Ugar Island. Map produced using Mapinfo (2019) with the Queensland basemap satellite imagery (2020).

Author Manuscript

This article is protected by copyright. All rights reserved

Figure S11: Locations of *Aedes albopictus* sweep net collections on St Pauls (Moa Island). Map produced using Mapinfo (2019) with the Queensland basemap satellite imagery (2020).

Figure S12: Locations of *Aedes albopictus* sweep net collections on Kubin (Moa Island). Map produced using Mapinfo (2019) with the Queensland basemap satellite imagery (2020).

Figure S13: Locations of *Aedes albopictus* sweep net collections on Warraber Island. Map produced using Mapinfo (2019) with the Queensland basemap satellite imagery (2020).

Figure S14: Location of sweep net sampling sites and detected incursions on Ngurapai between 2019-02-02 and 2019-02-09. Map produced using Mapinfo (2019) with the Queensland basemap satellite imagery (2020).

Figure S15: Individual ancestral lineage proportions for each village (n=22) with two ancestral lineages (K = 2) selected.

Figure S16: Individual ancestral lineage proportions for each village (n=22) with three ancestral lineages (K = 3) selected.

Figure S17: Individual ancestral lineage proportions for each village (n=22) with four ancestral lineages (K = 4) selected.

Figure S18: Scatter plot of a mantel test between geographic (natural log) and genetic distance (Rousset's a) for *Aedes albopictus* collections on Badu Island. Lines describe a linear regression fit with 95% Confidence Interval shown.

Figure S19: Scatter plot of a mantel test between geographic (natural log) and genetic distance (Rousset's a) for *Aedes albopictus* collections on Keriri Island. Lines describe a linear regression fit with 95% Confidence Interval shown.

Figure S20: Scatter plot of a mantel test between geographic (natural log) and genetic distance (Rousset's a) for *Aedes albopictus* collections on Masig Island in 2018. Lines describe a linear regression fit with 95% Confidence Interval shown.

Author Manuscript

This article is protected by copyright. All rights reserved

Figure S21: Scatter plot of a mantel test between geographic (natural log) and genetic distance (Rousset's a) for *Aedes albopictus* collections on Masig Island in 2019. Lines describe a linear regression fit with 95% Confidence Interval shown.

Figure S22: Scatter plot of a mantel test between geographic (natural log) and genetic distance (Rousset's a) for *Aedes albopictus* collections on Mer Island. Lines describe a linear regression fit with 95% Confidence Interval shown.

Figure S23: Scatter plot of a mantel test between geographic (natural log) and genetic distance (Rousset's a) for *Aedes albopictus* collections on Warraber Island. Lines describe a linear regression fit with 95% Confidence Interval shown.

Figure S24: Scatter plot of a mantel test between geographic (natural log) and genetic distance (Rousset's a) for *Aedes albopictus* collections on Erub Island. Lines describe a linear regression fit with 95% Confidence Interval shown.

Figure S25: Scatter plot of a mantel test between geographic (natural log) and genetic distance (Rousset's a) for *Aedes albopictus* collections on Mabuiag Island. Lines describe a linear regression fit with 95% Confidence Interval shown.

Figure S26: Scatter plot of a mantel test between geographic (natural log) and genetic distance (Rousset's a) for *Aedes albopictus* collections on Poruma Island. Lines describe a linear regression fit with 95% Confidence Interval shown.

Figure S27: Scatter plot of a mantel test between geographic (natural log) and genetic distance (Rousset's a) for *Aedes albopictus* collections on Iama Island. Lines describe a linear regression fit with 95% Confidence Interval shown.

Figure S28: Scatter plot of a mantel test between geographic (natural log) and genetic distance (Rousset's a) for *Aedes albopictus* collections on Ugar Island. Lines describe a linear regression fit with 95% Confidence Interval shown.

Author Manuscript

This article is protected by copyright. All rights reserved

Figure S29: Scatter plot of a mantel test between geographic (natural log) and genetic distance (Rousset's a) for *Aedes albopictus* collections on Kubin (Moa Island). Lines describe a linear regression fit with 95% Confidence Interval shown.

Figure S30: Scatter plot of a mantel test between geographic (natural log) and genetic distance (Rousset's a) for *Aedes albopictus* collections on St Pauls (Moa Island). Lines describe a linear regression fit with 95% Confidence Interval shown.

Figure S31: Full-size figure of Fig 3.

Figure S32: PCA-UMAP of all non-incursive genotypes, using the first 4 PCs, 50 neighbours and a 0.5 minimum distance.

Figure S33: Scree plot from pcadapt.

Figure S34: Genome-wide genetic structure of TSI villages, for the first and second PCs.

Figure S35: Genome-wide genetic structure of TSI villages, for the third and fourth PCs.

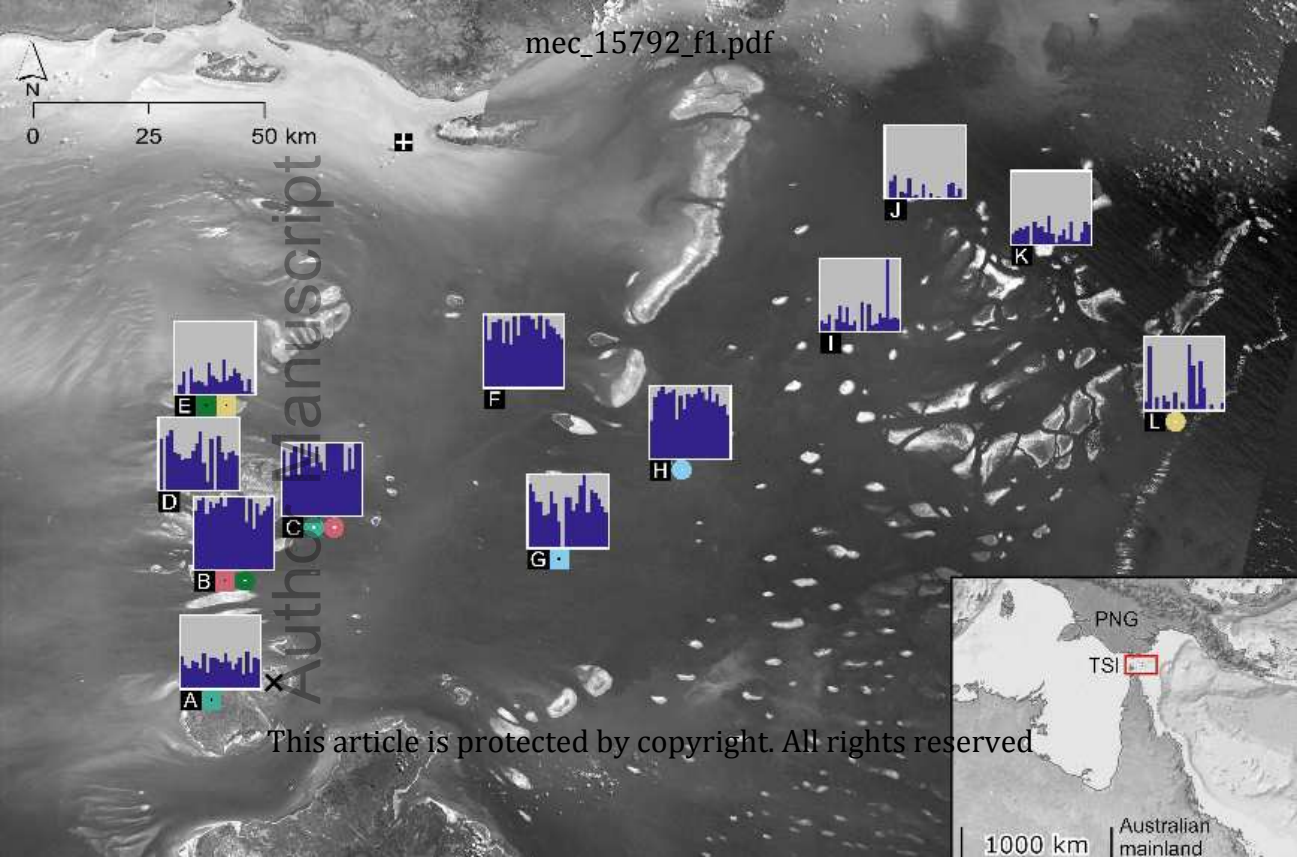
Figure S36: Genome-wide genetic structure of TSI villages, for the fifth and sixth PCs.

Figure S37: Genome-wide genetic structure of TSI villages, for the seventh and eigth PCs.

Author Manuscript

This article is protected by copyright. All rights reserved

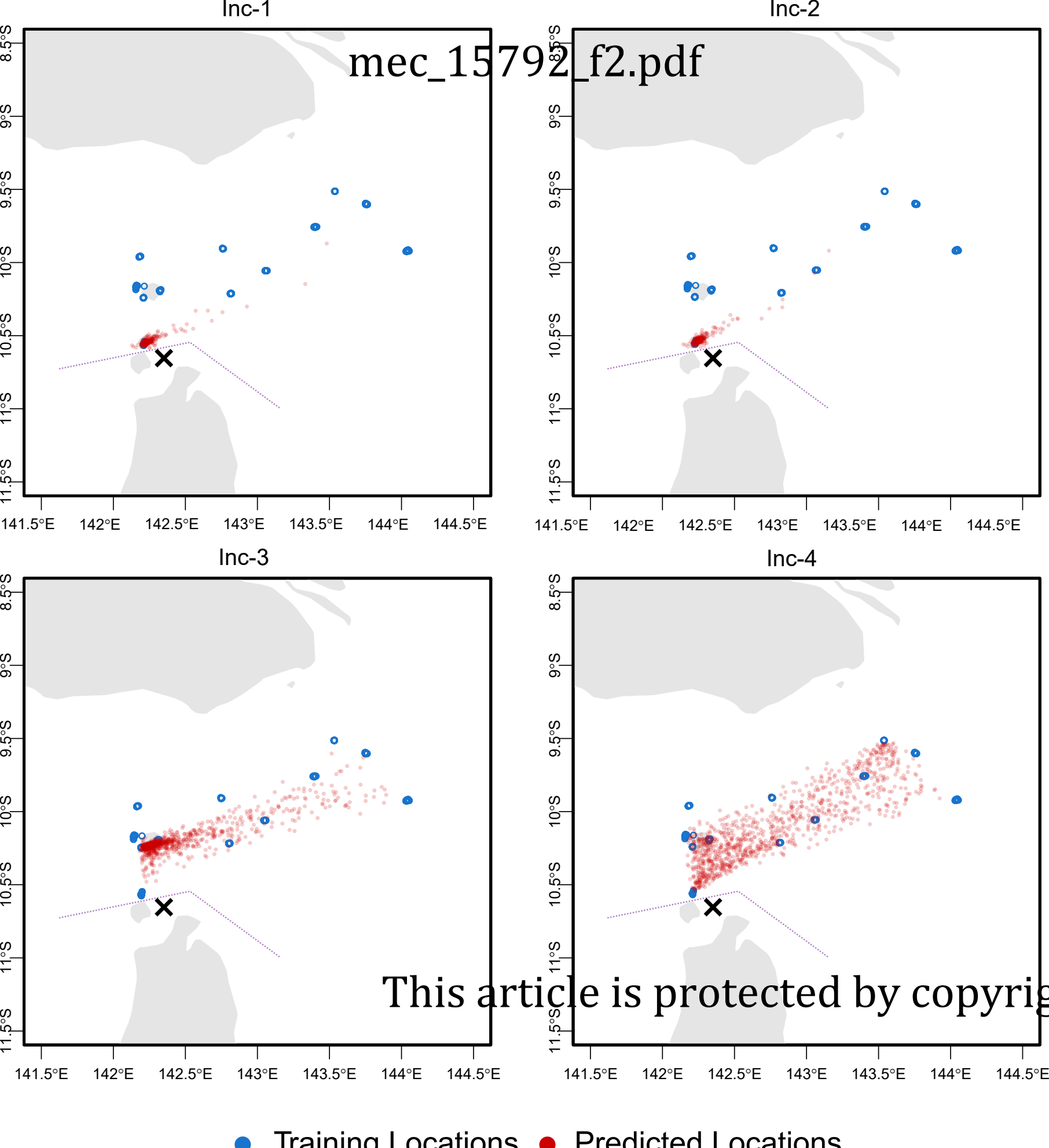
mec_15792_f1.pdf



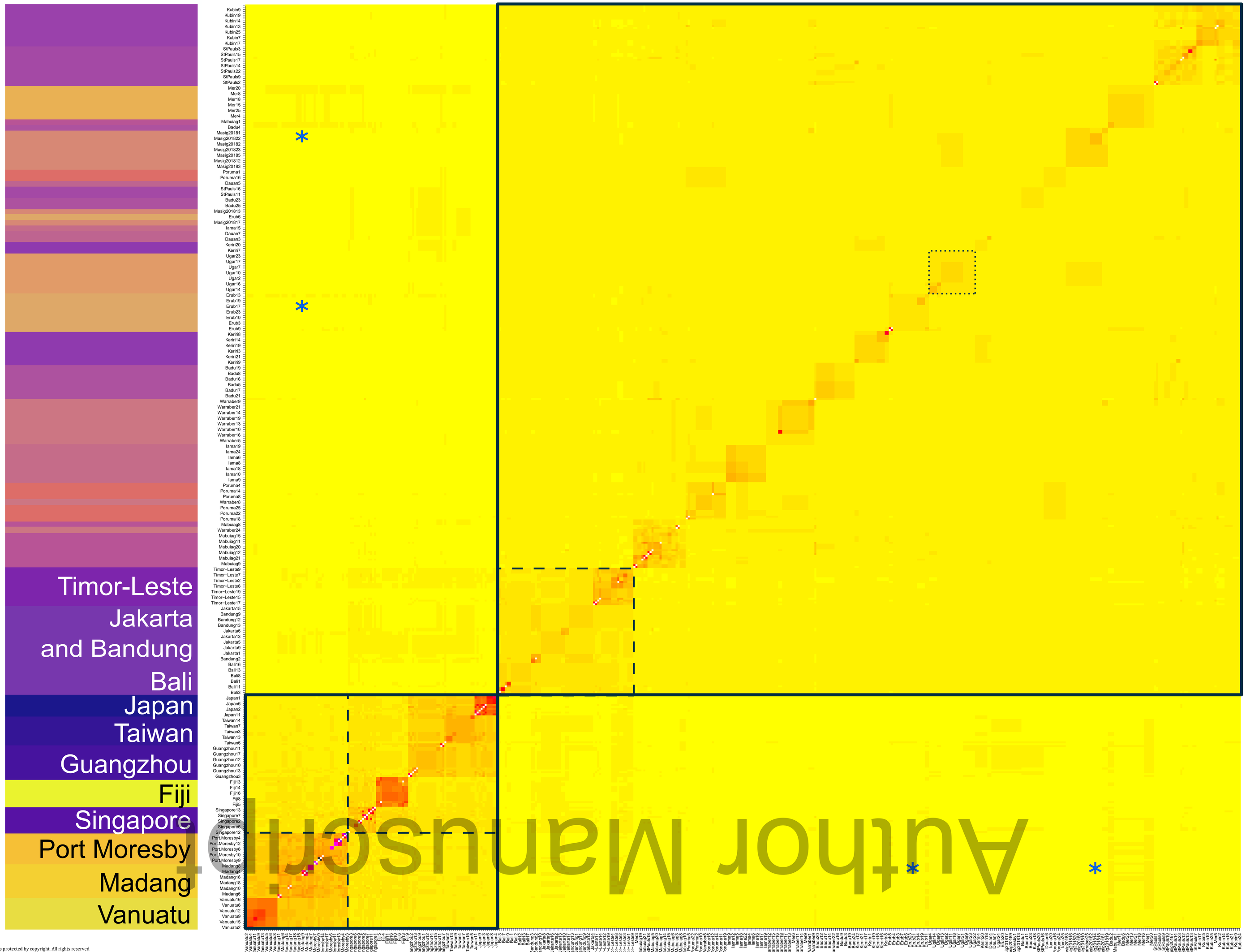
Author Manuscript

This article is protected by copyright. All rights reserved

mec_15792_f2.pdf



This article is protected by copyrig



This article is protected by copyright. All rights reserved