



Research paper

Viral genome-based Zika virus transmission dynamics in a paediatric cohort during the 2016 Nicaragua epidemic



Haoyang Sun^{a,1}, Raquel A. Binder^{b,c,1}, Borame Dickens^{a,1}, Paola Florez de Sessions^d,
Maia A. Rabaa^{e,f}, Eliza Xin Pei Ho^d, Alex R Cook^a, Fausto Bustos Carrillo^{b,c},
Jairo Carey Monterrey^c, Guillermina Kuan^{c,g}, Angel Balmaseda^{c,h}, Eng Eong Ooi^{a,i}, Eva Harris^{b,*},
October M. Sessions^{a,i,j,*}

^a Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore

^b Division of Infectious Diseases and Vaccinology, School of Public Health, University of California, Berkeley, CA, USA

^c Sustainable Sciences Institute, Managua, Nicaragua

^d Genome Institute of Singapore, A*STAR, Singapore, Singapore

^e The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam

^f Centre for Tropical Medicine, Nuffield Department of Medicine, Oxford University, Oxford, UK

^g Health Center Sócrates Flores Vivas, Ministry of Health, Managua, Nicaragua

^h Laboratorio Nacional de Virología, Centro Nacional de Diagnóstico y Referencia, Ministry of Health, Managua, Nicaragua

ⁱ Duke-NUS Medical School, Singapore, Singapore

^j Department of Pharmacy, National University of Singapore, Singapore

ARTICLE INFO

Article History:

Received 13 May 2021

Revised 2 August 2021

Accepted 9 September 2021

Available online 7 October 2021

Keywords:

Zika virus
transmission network
genomic inference
Nicaragua

ABSTRACT

Background: Nicaragua experienced a large Zika epidemic in 2016, with up to 50% of the population in Managua infected. With the domesticated *Aedes aegypti* mosquito as its vector, it is widely assumed that Zika virus transmission occurs within the household and/or via human mobility. We investigated these assumptions by using viral genomes to trace Zika transmission spatially.

Methods: We analysed serum samples from 119 paediatric Zika cases participating in the long-standing Paediatric Dengue Cohort Study in Managua, which was expanded to include Zika in 2015. An optimal spanning directed tree was constructed by minimizing the differences in viral sequence diversity composition between patient nodes, where low-frequency variants were used to increase the resolution of the inferred Zika outbreak dynamics.

Findings: Out of the 18 houses where pairwise difference in sample collection dates among all the household members was within 30 days, we only found two where viruses from individuals within the same household were up to 10th-most closely linked to each other genetically. We also identified a substantial number of transmission events involving long geographical distances (n=30), as well as potential super-spreading events in the estimated transmission tree.

Interpretation: Our finding highlights that community transmission, often involving long geographical distances, played a much more important role in epidemic spread than within-household transmission.

Funding: This study was supported by an NUS startup grant (OMS) and grants R01 AI099631 (AB), P01 AI106695 (EH), P01 AI106695-03S1 (FB), and U19 AI118610 (EH) from the US National Institutes of Health.

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

Zika virus (ZIKV) is a member of the enveloped *Flavivirus* genus, with a 10.7 kb positive-sense RNA genome, that is transmitted by *Aedes aegypti* and *Ae. albopictus* mosquitoes [1]. In addition to vector-

borne transmission, both perinatal and sexual transmission have been reported [2]. ZIKV was first discovered among rhesus monkeys of the Ziika Forest of Uganda in 1947, with the first human ZIKV isolate obtained in 1954 [3,4]. While serological data must be interpreted with caution due to cross-reactivity among flaviviruses, data suggest that ZIKV is endemic to Africa and Asia [5]. The first large Zika outbreak outside of Africa and Asia was reported on Yap Island (Western Pacific) in 2007 [6], and the second was occurred in French Polynesia, South Pacific, in 2013/14, where severe complications

* Corresponding authors.

E-mail address: October.sessions@nus.edu.sg (O.M. Sessions).

¹ Denotes equal contribution

Research in context

Evidence before this study

We searched PubMed for articles published from database inception to 30th June 2021 using the search terms ((Zika) OR (Dengue) OR (Chikungunya) OR (Yellow fever)) AND ((transmission tree) OR (transmission network)). We did not find any study utilizing whole-genome sequencing data to reconstruct transmission network for *Aedes*-borne disease outbreaks at a fine spatial scale.

Added value of this study

Our study used Zika virus genomic data collected from 119 participants of the long-standing Pediatric Dengue Cohort Study to estimate the Zika virus transmission network in Managua, Nicaragua. This added to the resolution of the inferred Zika virus transmission dynamics, and highlighted that community transmission, often involving long geographical distances, may have played a much more important role in epidemic spread than within-household transmission. Further, our results demonstrated the important contribution of transmission hotspots to the dissemination of the virus, as well as potential super-spreading events.

Implications of all the available evidence

Our study has emphasized the importance of implementing vector control measures outside as well as inside homes in order to successfully control future arboviral outbreaks in similar settings.

Health (NIC-MINSA/CNDR CIRE-09/03/07-008.ver14) approved the study protocol and the ongoing Nicaraguan PDCS. Written informed consent was collected from the parent or legal guardian of each child. In addition, children above 6 provided verbal assent.

Study Population

ZIKV-positive study participants and their respective households were drawn from the PDCS, a long-standing dengue cohort established in 2004 in Managua, Nicaragua. The study was expanded to include ZIKV infection starting in July 2015, with the first Zika case confirmed in January 2016. The PDCS study design, population, and detailed methods have been described previously [11, 14, 15]. Briefly, the study participants reside within the catchment area of the Health Center Sócrates Flores Vivas (HCSFV) in District II of Managua. The HCSFV serves 18 neighborhoods (barrios), mostly encompassing populations of low-to-middle socioeconomic status. Primary healthcare is provided by study personnel, and symptomatic ZIKV infections were identified by real-time reverse transcription PCR (rRT-PCR) in serum and/or urine and by serology (n=560) [11]. Zika cases included in this study were from the paediatric cohort and were all confirmed by rRT-PCR in serum (Table S1). We first included all the cases belonging to houses with two or more Zika cases. Among these 36 houses, 11 were identified where a serum sample with viral sequence was available for only one case. We then sampled 62 houses with only one rRT-PCR-confirmed Zika case. After further excluding 8 cases where sufficient sequencing data was not available to generate a consensus genome or call low-frequency variants, we obtained a total of 119 cases associated with complete genome or mostly complete sequence and epidemiological data for analysis, with their houses classified as follows:

- 1 Houses with ≥ 2 cases analysed: # houses = 24, # cases = 51;
- 2 Houses with only 1 case analysed: # houses = 68, # cases = 68.

Hereafter, we will refer to houses with ≥ 2 cases analysed as “indicator houses”. For each of the 119 patients, we also collected information on the school that he/she attended during the 2016 Zika epidemic (refer to Figure S1 for the geographic distribution of households and schools included in the analysis).

Viral sequencing

ZIKV genomes were sequenced from clinical samples according to the methods described in Kamaraj et al., 2019 [16]. Briefly, RNA was isolated from sera with Trizol (Life Technologies) according to the manufacturer's instructions. Illumina libraries were then constructed from total RNA using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs). Following library construction, ZIKV genomes were enriched using custom designed biotinylated, 120mer xGen Lockdown baits (Integrated DNA Technologies) complementary to the ZIKV genome. Enriched libraries were sequenced on an Illumina HiSeq 4000 (Genome Institute of Singapore). The quality of the reads generated were confirmed by FastQC [17] and Trim Galore [18] was used to trim and filter the reads with a minimum quality cutoff of 20 and a minimum read length of 35 nt. Initial mapping of the sequenced reads was performed with BWA mem aligner [19] against a contemporary ZIKV genome (Genbank ID KY765327.1). A consensus genome was then created using the bam2cons_iter.sh script from the ViPR pipeline [20] and a final mapping of the reads against this genome was used for subsequent variant analysis.

Transmission Network Reconstruction

To reconstruct the transmission network among the analysed serum samples from the paediatric cohort participants, we accounted

were first reported [7]. In 2015, ZIKV emerged in Brazil and subsequently spread rapidly across Latin America and the Caribbean region [8]. The World Health Organization declared Zika a Public Health Emergency of International Concern in February 2016 due to ongoing and widespread autochthonous ZIKV circulation and increasing evidence of severe ZIKV-associated complications, such as congenital birth defects, including microcephaly, in newborns and Guillain-Barré Syndrome in adults [8].

ZIKV was first detected in Nicaragua in January 2016, and the country experienced an explosive epidemic between June and September 2016 [9]. In a previous study conducted in Managua in 2017, ZIKV seroprevalence was found to be 36% and 56% in a paediatric cohort and an adult cohort, respectively [10]. Although estimates of the overall transmission intensity, its spatial variation, and the individual-level risk factors of infection exist [10, 11], our current understanding of the transmission pathways of ZIKV infections remains limited - key features of the epidemic are still unknown. Elsewhere, studies have mostly relied on inferring *Aedes*-borne disease transmission based on timing of onset of illness and geographical locations of identified cases [12, 13]. Here, we used ZIKV genomic data collected from 119 participants of the long-standing Pediatric Dengue Cohort Study (PDCS) to infer transmission dynamics of the 2016 Zika epidemic in Managua, Nicaragua. Specifically, we (i) estimated transmission networks at the individual and neighbourhood level; (ii) investigated the relative contributions of within- versus between-household transmission, as well as short-distance (< 1 km) versus long-distance (> 1 km) transmission to the epidemic; and (iii) identified potential super-spreading events.

Methods

Ethics

The Institutional Review Boards (IRB) of the University of California, Berkeley (#2010-09-2245), and the Nicaraguan Ministry of

for intra-host sequence diversity by incorporating low-frequency variants with depth of coverage above 100 into the calculation. Specifically, a consensus sequence of 10,807 nucleotides was first defined for the viral genome contents of each sample, where any gaps were filled with N's to normalize sequence length. Variants were then called using LoFreq 2 [21]. To visualize the evolutionary relationship among the 119 consensus sequences, we estimated the maximum likelihood phylogeny using the IQ-TREE [22], with the substitution model selected based on the Akaike Information Criterion [23] and branch supports assessed using the ultrafast bootstrap approximation [24].

At each position j of the viral genome in a serum sample i , we created a vector of length 6, storing the probabilities of observing adenine, cytosine, guanine, thymine, insertion, and deletion: $\mathbf{P}_{ij} = (p_{ij}^1, p_{ij}^2, p_{ij}^3, p_{ij}^4, p_{ij}^5, p_{ij}^6)$. For each pair of serum samples i_1 and i_2 (from two different individuals), we quantified the difference in the viral sequence diversity composition via calculating the genetic distance $d_{\text{Genetic}}(i_1, i_2)$ below, which is bounded between 0 and 1:

$$d_{\text{Genetic}}(i_1, i_2) = \frac{1}{L} \sum_{j=1}^L \left(\frac{1}{\sqrt{2}} \cdot \|\sqrt{\mathbf{P}_{i_1 j}} - \sqrt{\mathbf{P}_{i_2 j}}\|_2 \right) \quad (1)$$

In Equation 1, the Hellinger distance between the two probability distributions $\mathbf{P}_{i_1 j}$ and $\mathbf{P}_{i_2 j}$ at each position j was computed [25] and subsequently averaged across the entire genome (with length L) to measure the genetic distance between viral populations in serum samples i_1 and i_2 . Incorporation of low-frequency variants into the analysis yields more information to be used for transmission network reconstruction compared with relying on consensus sequences alone. To demonstrate this, we compared the coefficients of variation of (i) the pairwise genetic distances derived using Eq. 1 with (ii) the pairwise p-distances based on the consensus sequences only.

The genetic distance matrix computed following Eq. 1 was first fed into the “gengraph” function in the R package “adeget” [26], to determine the optimal number of distinct clusters. Should this value exceed one, transmission network reconstruction would need to be performed separately for cases within each cluster. Estimation of the transmission tree was performed using the SeqTrack algorithm, which minimizes the total genetic distances along all of the edges between patient nodes, subject to the constraint that the sample collection date of an ancestor must precede that of its descendent [27]. We did not impose any geographical distance constraints for patients with a direct ancestor-descendent relationship during the estimation of the final tree, as the geographical distance between any two houses included within the study was reasonably short (median 0.99 km, IQR: 0.63 km–1.40 km). Instead, a sensitivity analysis was carried out to quantify the robustness of our results (i.e., to assess the extent to which each estimated transmission link is supported by the genomic data), where we imposed a higher penalty for pairs with a greater geographical distance, controlling for the length of time between the two sample collection dates. Specifically, we generated 100 equally spaced values for the penalty coefficient, α , ranging from zero to a pre-defined maximum value, α_{max} , under which the genetic distance and the penalty term shared an equal range. In Eq. 2 below, $D(i_1, i_2)$ represents the “total distance” between samples i_1 and i_2 , which was used to construct a transmission tree in the sensitivity analysis. Its first term is the genetic distance between the two samples as defined previously, and the second term is the product of the penalty coefficient and the underlying average dissemination speed of the virus if one patient was indeed the most recent sampled ancestor of the other. Here, we used the length of interval between the two sample collection dates, denoted by $|t_{i_1} - t_{i_2}|$, to approximate the time difference between the two infections, and the geographical distance between the two home locations, $d_{\text{Geographic}}(i_1, i_2)$, was approximated using the Haversine function.

$$D(i_1, i_2) = d_{\text{Genetic}}(i_1, i_2) + \alpha \cdot \frac{d_{\text{Geographic}}(i_1, i_2)}{|t_{i_1} - t_{i_2}|} \quad (2)$$

For each link in our final results (corresponding to the $\alpha = 0$ scenario), we computed the proportion of times the link appeared in the transmission trees as we gradually increased the penalty coefficient to the maximum value. This served as a robustness measure to highlight long-distance transmission links with reasonably high confidence. Robust edges corresponding to a difference in sample collection dates (hereafter referred to as edge time length) within the range of 10 to 30 days would be of particular interest, since they may represent “first-generation transmission events” (i.e., host-mosquito-host transmission) given that the serial interval estimates obtained by most studies were found to fall within this range [28]. We did not exclude estimated transmission links with edge time lengths shorter than 10 days, as some of these links could still represent transmission events in reality due to the variation in the time between dates of symptom onset and sample collection.

To obtain further insights into the transmission dynamics among the paediatric cohort participants, for each study member in an indicator house, we identified whether the individual having the lowest genetic distance to that household member belonged to the same household, where we restricted ourselves to all households where the pairwise difference in sample collection dates among all the household members was within 30 days. In addition, we also counted the number of estimated first-generation transmission events that involved children from the same school. To identify potential super-spreading events, we calculated the total number of first-generation transmission events for each study participant that he/she had seeded as estimated using the SeqTrack algorithm (hereafter referred to as the degree of each node). Super-spreading events were detected by identifying nodes whose degrees exceeded the 90th percentile, where we restricted ourselves to all nodes with at least one offspring as estimated by the model.

The analyses described so far were based on the genetic distances that accounted for intra-host sequence diversity. To compare and contrast results, we additionally repeated all the aforementioned analyses using the p-distance calculated based on the consensus sequences only. All statistical analyses were performed using R version 3.5.3 [29].

Role of funding source

The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Serum samples from 119 rRT-PCR-confirmed Zika cases belonging to 92 households of the PDCS were analysed. Among these participants, 61 (51.3%) were male and 58 (48.7%) were female. The age of the paediatric participants ranged from 2 to 14 years, with 24 (20.2%) being between 2 and 5 years old, 37 (31.1%) between 6 and 9 years old, and 58 (48.7%) between 10 and 14 years old. The vast majority of serum samples were collected between July and September 2016 (Fig. 1a), and the pairwise geographical distances among all the study participants were within the range of 0 km–3 km (Fig. 1b). Notably, incorporating low-frequency variants into our analyses helped to increase the variability of the pairwise genetic distances compared with using consensus sequences alone (Figs. 1c, 1d), as evidenced by a 56% increase in the coefficient of variation of the genetic distances, from 0.351 to 0.549. This allowed for a more detailed and accurate transmission tree reconstruction. In addition, the number of clusters

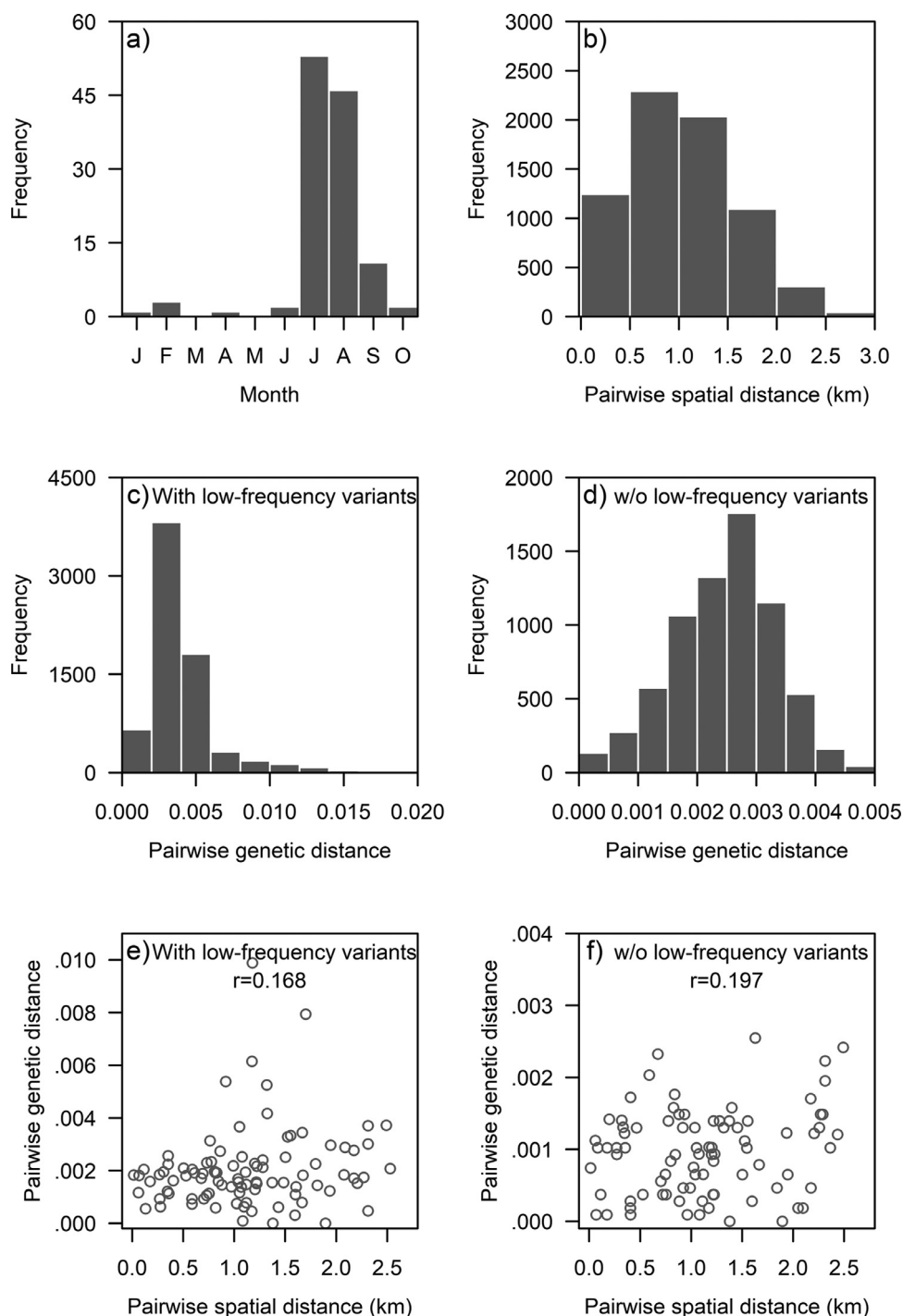
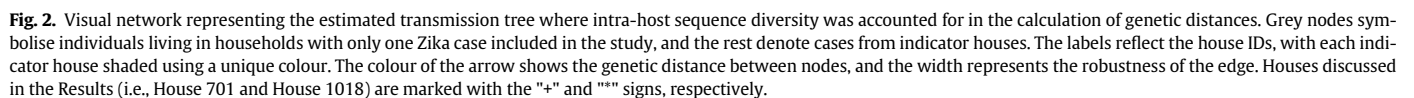


Fig. 1. Exploratory analysis of input data. (a) Histogram of ZIKV-positive serum sample collection dates; (b) Histogram of pairwise geographical distances among all the 119 paediatric cohort participants analysed; (c, d) Histograms of pairwise genetic distances computed with and without intra-host sequence diversity accounted for, respectively; (e, f) Scatterplots of pairwise genetic distance versus the pairwise geographical distance, for all estimated transmission links with edge time lengths within 30 days, with and without low-frequency variants, respectively.

estimated by “gengraph” was one regardless of the genetic distance measure used; hence, we reconstructed a single transmission tree linking all the cases included in our analyses. Unless otherwise stated, all the results presented below were derived from the analyses where intra-host sequence diversity was accounted for.

A visual network representing our estimated transmission tree (Fig. 2) was assembled where each indicator house was shaded with a unique colour and grey nodes symbolise individuals living in households with one Zika case included in our analyses. Interestingly, only a few indicator houses contained patients that shared a common

most recent sampled ancestor or who were directly linked in the estimated transmission tree ($n=7$, e.g., House 1018). Instead, many indicator houses contained Zika cases whose most recent sampled ancestors came from different houses ($n=17$, e.g., House 701). Moreover, out of the 18 indicator houses whose household members had sample collection dates ranging within 30 days, we only found two houses (11%) where viruses from individuals within each house had the lowest genetic distance to each other. When we relaxed the condition by counting the number of indicator houses where viruses from individuals within the same household were at least 10th-most



We observed a predominant transmission wave from west to east prior to mid-July, with many first-generation transmission events estimated to be seeded from the Cuba barrio (Figs. 3a–3b), followed by cross-neighbourhood epidemic spread in various directions (Figs. 3c–3d) based on the reconstructed transmission tree (refer to Video S1 for a continuous visualization of all the estimated

transmission events). The median difference in sample collection dates for all ancestor-descendent pairs in the estimated transmission tree was 14 days (IQR: 6–28), with 39% of the edge time lengths falling within the range of 10 to 30 days (Fig. 4a), which may correspond to first-generation transmission events (Table S2). The majority of the edges were highly robust even when we imposed a large penalty on the geographical distance (Fig. 4b). In particular, out of all the inferred first-generation transmission events involving a long geographical distance (>1 km), 50% were found to have a robustness value of one. More than 50% of the inferred first-generation transmission events involved long geographical distances (> 1 km), and this percentage remained above 50% even if we restricted ourselves to all first-generation transmission events with robustness values being 1.

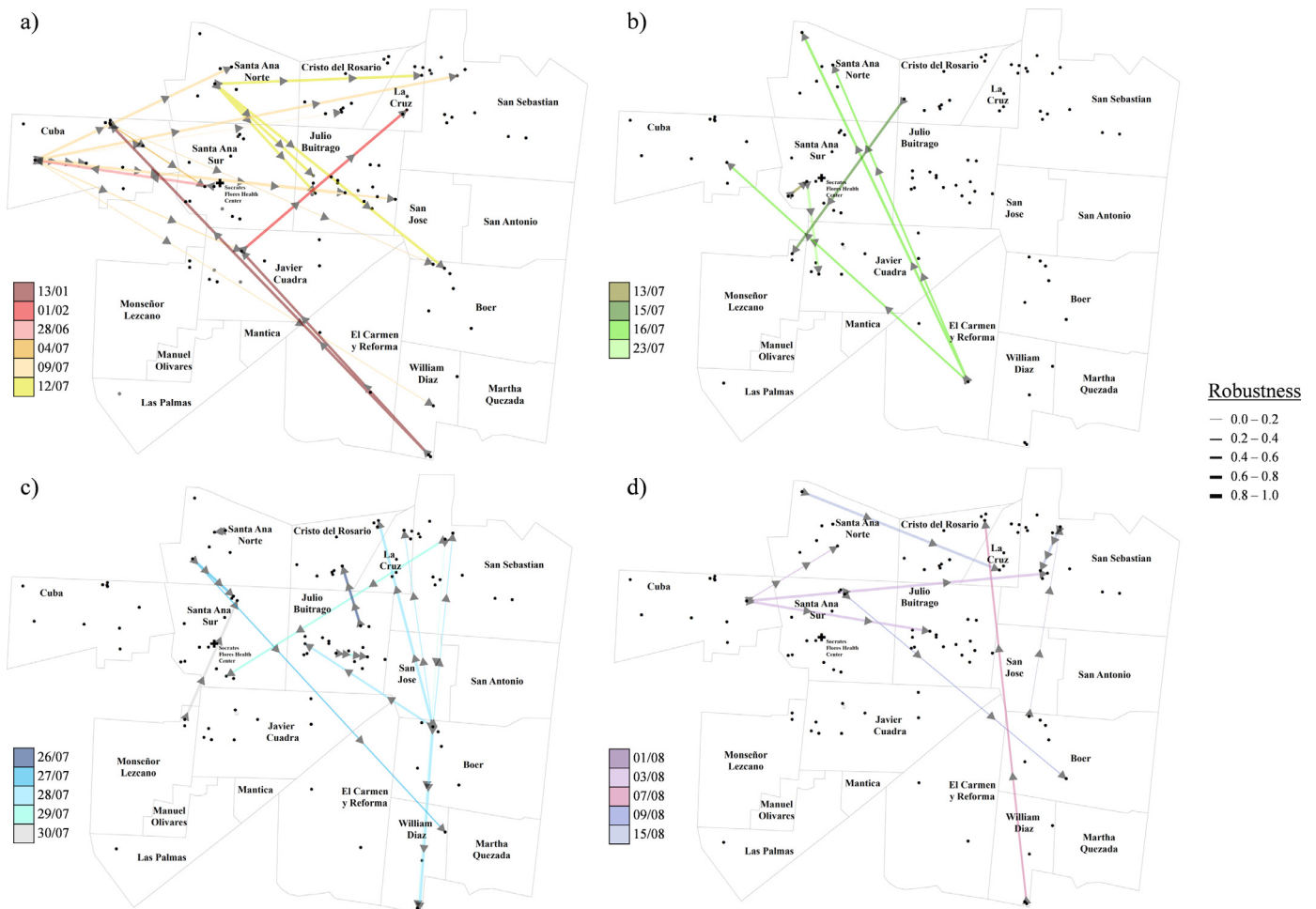


Fig. 3. Visualization of all the inferred first-generation transmission events in the analysis, where intra-host sequence diversity was accounted for in the calculation of genetic distances. The colours indicate the sample collection dates of each inferred ancestor case. Predicted first-generation transmission events through time in 2016 are presented in panel (a) from 13 January to 12 July, (b) from 13 July to 23 July, (c) from 26 July to 30 July, and (d) from 1 August to 15 August. Refer to Video S1 for the continuous visualization of all the inferred transmission events in the study.

Overall, the pairwise geographical distance and pairwise genetic distance were found to have a weak positive correlation (Figs. 1e, 1f), regardless of whether intra-host sequence diversity was accounted for.

Interestingly, out of all the estimated first-generation transmission events, 13 (28%) were seeded from the Cuba barrio (Fig. 5a). Notably, we identified 3 study participants from Santa Ana Norte,

Boer, and Cuba respectively who were estimated to have seeded super-spreading events, with the degrees of their nodes being 4, 5, and 8. As we started to impose penalties on geographical distances (i.e., including the second term of Eq. 2 in the total distance calculation), cases in Santa Ana Norte and Santa Ana Sur were found to have an increased contribution to the epidemic spread. Nonetheless, cases

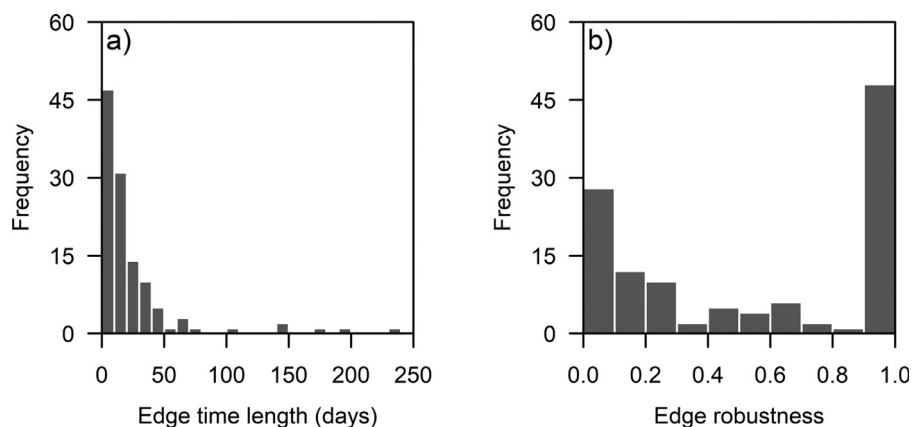


Fig. 4. Summary statistics of the model output where intra-host sequence diversity was accounted for in the calculation of genetic distances. (a) Histogram of the difference in sample collection dates between each individual and its most recent sampled ancestor in the estimated transmission tree; (b) Histogram of robustness value for each edge in the estimated transmission tree as penalties were imposed on geographical distance controlling for the edge time length.

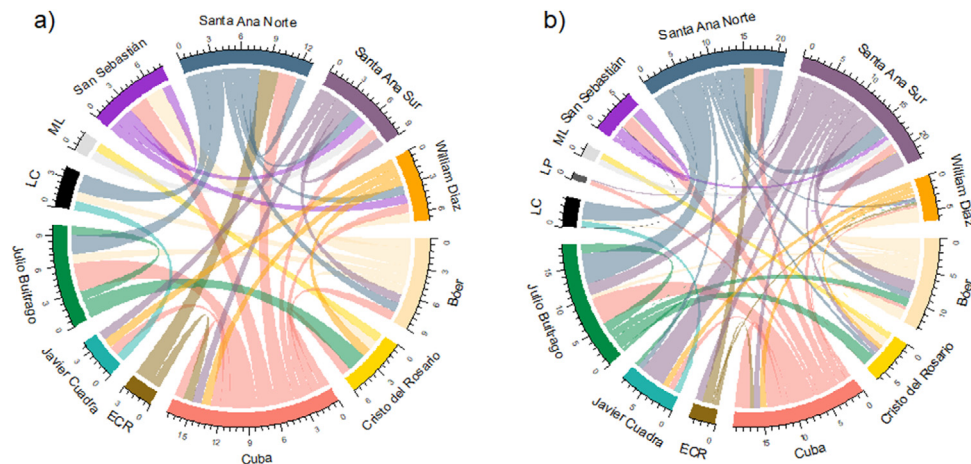


Fig. 5. Chord diagram visualizing inferred first-generation transmission events between neighbourhoods, where intra-host sequence diversity was accounted for in the calculation of genetic distances. The colour of each line indicates the origin neighbourhood. (a) No penalty imposed on geographical distance (corresponding to the main analysis), and (b) results averaged over all the 100 trees as the penalty coefficient was varied. Overall, penalizing links with large geographical distances only caused a modest change in the transmission network estimated at the neighborhood level. ECR: El Carmen y Reforma, LC: La Cruz, LP: Las Palmas, ML: Monseñor Lezcano.

in barrio Cuba remained one of the largest contributors to the inferred first-generation transmission events based on the results averaged over all 100 trees as we varied the penalties on geographical distances (Fig. 5b). Overall, penalizing links with large geographical distances only caused a modest change in the estimated transmission network at the neighbourhood level, indicating that our results were robust and that the genetic data provide strong support for long-distance transmission. Moreover, despite 46% of the inferred individual transmission links being different between the two transmission trees reconstructed using different genetic distance measures, the general dynamics of the epidemic (e.g., important contribution of cases in Cuba and long-distance transmission to the epidemic spread) remained similar (Table S3). All the results obtained based on the consensus sequences, including the estimated transmission tree (Fig. S2), visualization of all the inferred first-generation transmission events (Fig. S3), summary statistics of the model output (Fig. S4), Chord diagram visualizing the transmission network estimated at the neighbourhood level (Fig. S5), and the phylogenetic tree (Fig. S6) can be found in the supporting information.

Discussion

Understanding the spatial spread of infectious diseases is critical for optimizing resource allocation for timely outbreak management, and there has been a growing body of research that attempts to utilize outbreak data to design intervention strategies. In a spatiotemporal modelling study conducted by Guzzetta *et al.* in Porto Alegre, Brazil, around 70% of dengue virus transmission events were estimated to have occurred between individuals (via *Aedes* mosquitoes) whose residential locations were within 500 m from each other. This suggests that in this cohort, short-distance movement (<1 km) could explain the vast majority of dengue cases [30]. On average, the authors estimated that clusters of dengue transmission expanded slowly at a rate of ~600 m per month [30]. In contrast, our results highlight the contribution of long-distance ZIKV transmission to epidemic spread. The differences in the inferred spatial dynamics between the two studies can be in part explained by the input data and model specifications. Specifically, Guzzetta *et al.* utilized an exponential spatial kernel to regulate the transmission probability between an infector and infectee based on the distance between their home locations [30]. Although this performed much better than a radiation model representing long-distance transmission [30], inclusion of additional information, such as non-residential exposure

locations, may uncover important transmission links involving much longer distances [31]. To allow for a more accurate assessment of the Zika epidemic dynamics in Managua, we used viral genome data to estimate the transmission network among the paediatric cohort participants, and the insights we found are unlikely to be obtained from models solely based on home location and time of symptom onset. Despite the limited flight range of *Aedes* mosquitoes, our study has identified long-distance transmission to play an important role in the epidemic spread, which can be possibly explained by the observed human movement patterns within the study area, where children moved around Managua at a relatively high frequency, including to the neighbouring barrios in District II.

Our findings are generally congruent with other studies aiming to understand arboviral epidemics. For example, it was estimated that within-household transmission only accounted for 22% of all ZIKV infections detected during an outbreak in Martinique in 2016, using data collected from a household transmission study, blood-donor seroprevalence studies, and laboratory-testing results among pregnant women with Zika-like illness [32]. Spatiotemporal modelling of data collected from the 2008–2009 dengue epidemic in Cairns, Australia, suggested that more than 50% of potential exposure locations were non-residential [33]. Within our study, the Julio Buitrago neighbourhood experienced multiple cases relatively clustered around a large playground area, suggesting locations where children regularly congregate during the day may require heightened vector control to prevent between-household transmission. There are also multiple tire shops near the playground, which could also explain the cluster of cases, as tires regularly fill with water and become breeding grounds for *Aedes* mosquitoes. Our findings, however, do not imply that protecting individuals from arbovirus transmission at home is unnecessary, since between-household transmission could also occur during house visits, and within-household transmission may still play a non-negligible role in epidemic spread. For instance, a 2008 contact cluster study of 2,444 individuals in Iquitos, Peru, found that house-to-house human movements underlie patterns of infection and contribute to both temporal and spatial heterogeneity in dengue incidence [34]. In another study in Puerto Rico, households with open windows and doors had a significantly higher chance of ZIKV infection during the 2016–2017 epidemic [35], highlighting the importance of household-based interventions in addition to disease control activities targeted at non-residential locations to reduce ZIKV and other arbovirus transmission.

Dengue virus and chikungunya virus have similar transmission cycles and vectors as ZIKV. As a result, our results regarding the low within-household transmission of ZIKV may also extend to other arboviral epidemics in this population, particularly the two large chikungunya epidemics in 2014 and 2015 since they directly preceded the 2016 Zika epidemic, and factors favouring transmission are unlikely to have changed dramatically in a short time period. In fact, a parallel spatial study of these same chikungunya and Zika epidemics found a low intra-cluster correlation coefficient of infections within households [36]. However, our results may not be generalizable to other study settings, as transmission is often context-specific.

We have identified the Cuba neighbourhood to be an important hotspot for the widespread distribution of Zika cases in this outbreak, and this was evidenced by the observation that 13 out of the 46 inferred first-generation transmission events were seeded from barrio Cuba, with one case residing in the Cuba neighbourhood estimated to have seeded eight first-generation transmission events. The phenomenon that a small number of individuals are responsible for a disproportionately large number of transmission events has been recorded in numerous modelling studies [37], and additional evidence is needed to further our understanding of the underlying factors shaping these potential super-spreading events. Entomological surveillance can be used to identify the presence of especially productive vector habitats: for example, Padmanabha *et al.* found that 92% of *Ae. aegypti* pupae were located in only 5% of houses based on census data collected in Armenia, Colombia, during 2007–2009 [38]. In our study area, it has been proposed that the Central Cemetery of Managua, located next to barrio Cuba, may have served as an important mosquito-breeding site to amplify ZIKV spread [10,36]. In addition, the Cuba barrio houses a recycling centre where residents of surrounding barrios bring materials for recycling as well as a popular sports complex, which could also potentially account for human mobility and concentration in barrio Cuba, and these hypotheses need to be further examined when higher-resolution human movement data become available. It is worth noting, however, that as we imposed penalties on geographical distance, Santa Ana Sur and Santa Ana Norte were found to have an increased contribution to epidemic spread compared with the zero penalty scenario, indicating some residual uncertainty in the estimated transmission network at the neighbourhood level.

Our study has several limitations. Given the 37.8% of ZIKV infections that were estimated to be symptomatic in our study area [39] and the lack of adult cases in our paediatric study, the true underlying transmission network can only be partially revealed, as both inapparent infections and adult cases could be important contributors to the epidemic spread. Despite imperfect sampling, to our knowledge, we have obtained one of the most densely sampled whole-genome sequence datasets in a Zika outbreak setting. The overall spatiotemporal pattern of the epidemic inferred at the barrio level, as well as the discovery of potential super-spreading events, is presumed to be unlikely to change drastically if the sampling rate had been higher. However, if all the infected individuals from the households had been included in our study, the estimated relative contribution of within- and between-household transmission to the epidemic could potentially be somewhat modified. In addition, the paucity of information related to each patient's daily mobility hinders effective investigation of important exposure locations and further refinement of our tree estimation. Though we collected and analyzed information on the school attended by each patient during 2016, we only found one estimated first-generation transmission event involving two children from the same school and were unable to generate sufficient evidence to pinpoint where vector control should be targeted outside the household. Thus, given the relatively restricted size of the study area and only the residential and school location of each patient available, we did not impose distance-based penalties when we reconstructed the final transmission tree. A separate sensitivity

analysis was carried out instead, which showed that the estimated transmission network at the neighbourhood level was generally robust in the presence of geographical distance penalties. Unfortunately, data on the viral sequences from mosquitoes was not available, which could have helped to link the Zika cases to spatial locations where transmission may have occurred, thereby providing further insights into the ZIKV transmission dynamics within our paediatric cohort.

In conclusion, we have used whole-genome sequence data to infer the transmission dynamics of the 2016 Zika epidemic in a paediatric cohort within District II of Managua. We found that community transmission, often involving long geographical distances, played a much more important role in epidemic spread than within-household transmission. We also discovered potential super-spreading events that require further investigation. Our study has emphasized the importance of implementing vector control measures outside as well as inside homes in order to successfully control future arboviral outbreaks in similar settings.

Contributors

HS, RAB, BD, EEO, EH and OMS conceived the experiments and wrote the initial draft of the paper. PFS, MAR, EXPH and OMS prepared samples for sequencing and conducted the genome assemblies. HS, RB, BD, ARC, FBC performed the geospatial and genomic clustering analyses. RAB, JCM, GK, AB and EH coordinated the cohort and collected the samples. Underlying data was verified by HS, RAB, BD, EH and OMS. All authors read and approved the final version of the manuscript.

Data Sharing Statement

The viral sequence data supporting the findings of this study are available in GenBank (GenBank IDs: OK054369 - OK054487).

Declaration of Competing Interest

The authors declare no potential conflict of interests related to this work.

Acknowledgements

This study was supported by an NUS start-up grant to OMS and grants R01 AI099631 (AB), P01 AI106695 (EH), P01 AI106695-03S1 (FB), and U19 AI118610 (EH) from the National Institute of Allergy and Infectious Diseases of the US National Institutes of Health. ARC and HS are supported by the Singapore Ministry of Health's National Medical Research Council under the Centre Grant Programme - Singapore Population Health Improvement Centre (NMRC/CG/C026/2017_NUHS). We would also like to acknowledge Tuan Nini for her artwork related to the study.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ebiom.2021.103596](https://doi.org/10.1016/j.ebiom.2021.103596).

References

- [1] Musso D, Gubler DJ. Zika Virus. *Clin Microbiol Rev* 2016;29(3):487–524.
- [2] Oster AM, Brooks JT, Stryker JE, Kachur RE, Mead P, Pesik NT, et al. Interim Guidelines for Prevention of Sexual Transmission of Zika Virus - United States, 2016. *MMWR Morb Mortal Wkly Rep* 2016;65(5):120–1.
- [3] Dick GW, Kitchen SF, Haddock AJ. Zika virus. I. Isolations and serological specificity. *Trans R Soc Trop Med Hyg* 1952;46(5):509–20.
- [4] Macnamara FN. Zika virus: a report on three cases of human infection during an epidemic of jaundice in Nigeria. *Trans R Soc Trop Med Hyg* 1954;48(2):139–45.

- [5] Faye O, Freire CC, Iamarino A, Faye O, de Oliveira JV, Diallo M, et al. Molecular evolution of Zika virus during its emergence in the 20(th) century. *PLoS Negl Trop Dis* 2014;8(1):e2636.
- [6] Duffy MR, Chen TH, Hancock WT, Powers AM, Kool JL, Lanciotti RS, et al. Zika virus outbreak on Yap Island, Federated States of Micronesia. *N Engl J Med* 2009;360(24):2536–43.
- [7] Musso D, Bossin H, Mallet HP, Besnard M, Brout J, Baudouin L, et al. Zika virus in French Polynesia 2013–14: anatomy of a completed outbreak. *Lancet Infect Dis* 2018;18(5):e172–e82.
- [8] Kindhauser MK, Allen T, Frank V, Santhana RS, Dye C. Zika: the origin and spread of a mosquito-borne virus. *Bull World Health Organ* 2016;94(9):675–86C.
- [9] PAHO/WHO. Zika: Epidemiological Report: Nicaragua. 2017 2017 [cited 2017 25 September]. Available from: <https://www.paho.org/hq/dmdocuments/2017/2017-phe-zika-situation-report-nic.pdf>.
- [10] Zambrana JV, Bustos Carrillo F, Burger-Calderon R, Collado D, Sanchez N, Ojeda S, et al. Seroprevalence, risk factor, and spatial analyses of Zika virus infection after the 2016 epidemic in Managua, Nicaragua. *Proc Natl Acad Sci U S A*. 2018;115(37):9294–9.
- [11] Burger-Calderon R, Bustos Carrillo F, Gresh L, Ojeda S, Sanchez N, Plazaola M, et al. Age-dependent manifestations and case definitions of paediatric Zika: a prospective cohort study. *Lancet Infect Dis* 2020;20(3):371–80.
- [12] Aldstadt J, Yoon IK, Tannitsupawong D, Jarman RG, Thomas SJ, Gibbons RV, et al. Space-time analysis of hospitalised dengue patients in rural Thailand reveals important temporal intervals in the pattern of dengue virus transmission. *Trop Med Int Health* 2012;17(9):1076–85.
- [13] Sedda L, Vilela APP, Aguiar E, Gaspar CHP, Goncalves ANA, Olmo RP, et al. The spatial and temporal scales of local dengue virus transmission in natural settings: a retrospective analysis. *Parasit Vectors* 2018;11(1):79.
- [14] Gordon A, Kuan G, Mercado JC, Gresh L, Aviles W, Balmaseda A, et al. The Nicaraguan pediatric dengue cohort study: incidence of inapparent and symptomatic dengue virus infections, 2004–2010. *PLoS Negl Trop Dis* 2013;7(9):e2462.
- [15] Kuan G, Gordon A, Aviles W, Ortega O, Hammond SN, Elizondo D, et al. The Nicaraguan pediatric dengue cohort study: study design, methods, use of information technology, and extension to other infectious diseases. *Am J Epidemiol* 2009;170(1):120–9.
- [16] Kamaraj US, Tan JH, Xin Mei O, Pan L, Chawla T, Uehara A, et al. Application of a targeted-enrichment methodology for full-genome sequencing of Dengue 1–4, Chikungunya and Zika viruses directly from patient samples. *PLoS Negl Trop Dis* 2019;13(4):e0007184.
- [17] Andrews S. FastQC: A quality control tool for high throughput sequence data [Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>].
- [18] Krueger F. Trim Galore [Available from: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/].
- [19] Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26(5):589–95.
- [20] Wilm A. ViPR pipeline [Available from: <https://github.com/CSB5/vipr>].
- [21] Wilm A, Aw PP, Bertrand D, Yeo GH, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res* 2012;40(22):11189–201.
- [22] Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32(1):268–74.
- [23] Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14(6):587–9.
- [24] Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 2018;35(2):518–22.
- [25] Hellinger E. Neue Begründung der Theorie quadratischer Formen von unendlich-vielen Veränderlichen. *Crelles Journal* 1909;1909(136):210–71.
- [26] Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 2008;24(11):1403–5.
- [27] Jombart T, Eggo RM, Dodd PJ, Balloux F. Reconstructing disease outbreaks from genetic data: a graph approach. *Heredity (Edinb)* 2011;106(2):383–90.
- [28] Keegan LT, Lessler J, Johansson MA. Quantifying Zika: Advancing the Epidemiology of Zika With Quantitative Models. *J Infect Dis* 2017;216(suppl_10):S884–S90.
- [29] R_Core_Team. The R Project for Statistical Computing 2020 [Available from: <https://www.r-project.org>].
- [30] Guzzetta G, Marques-Toledo CA, Rosa R, Teixeira M, Merler S. Quantifying the spatial spread of dengue in a non-endemic Brazilian metropolis via transmission chain reconstruction. *Nat Commun* 2018;9(1):2837.
- [31] Prem K, Lau MSY, Tam CC, Ho MZJ, Ng LC, Cook AR. Inferring who-infected-whom-where in the 2016 Zika outbreak in Singapore—a spatio-temporal model. *J R Soc Interface* 2019;16(155):20180604.
- [32] Cousien A, Abel S, Monthieux A, Andronico A, Calmont I, Cervantes M, et al. Assessing Zika Virus Transmission Within Households During an Outbreak in Martinique, 2015–2016. *Am J Epidemiol* 2019;188(7):1389–96.
- [33] Vazquez-Prokopec GM, Montgomery BL, Horne P, Clennon JA, Ritchie SA. Combining contact tracing with targeted indoor residual spraying significantly reduces dengue transmission. *Sci Adv* 2017;3(2):e1602024.
- [34] Stoddard ST, Forshey BM, Morrison AC, Paz-Soldan VA, Vazquez-Prokopec GM, Astete H, et al. House-to-house human movement drives dengue virus transmission. *Proc Natl Acad Sci U S A*. 2013;110(3):994–9.
- [35] Rosenberg ES, Doyle K, Munoz-Jordan JL, Klein L, Adams L, Lozier M, et al. Prevalence and Incidence of Zika Virus Infection Among Household Contacts of Patients With Zika Virus Disease, Puerto Rico, 2016–2017. *J Infect Dis* 2019;220(6):932–9.
- [36] Carrillo FABM, B.L.; Monterrey, J.C.; Collado, D.; Saborio, S.; Miranda, T.; Barilla, C.; Ojeda, S.; Sanchez, N.; Plazo, M.; Laguna, H.S.; Elizondo, D.; Arguello, S.; Gajewski, A.M.; Maier, H.E.; Latta, K.; Carlson, B.; Coloma, J.; Katzelnick, L.; Sturrock, H.; Balmaseda, A.; Kuan, G.; Gordon, A.; Harris, E. Epidemics of chikungunya, Zika, and COVID-19 reveal bias in case-based mapping. 2021.
- [37] Stein RA. Super-spreaders in infectious diseases. *Int J Infect Dis* 2011;15(8):e510–3.
- [38] Padmanabha H, Durham D, Correa F, Diuk-Wasser M, Galvani A. The interactive roles of Aedes aegypti super-production and human density in dengue transmission. *PLoS Negl Trop Dis* 2012;6(8):e1799.
- [39] Burger-Calderon R, Gonzalez K, Ojeda S, Zambrana JV, Sanchez N, Cerpas Cruz C, et al. Zika virus infection in Nicaraguan households. *PLoS Negl Trop Dis* 2018;12(5):e0006518.