Multiple introductions of Zika virus into the United States revealed through genomic epidemiology

Nathan D Grubaugh[1*], Jason T Ladner[2,*], Moritz UG Kraemer[3,*], Gytis Dudas[4,*], Amanda L Tan[5,*], Karthik Gangavarapu[1*], Michael R Wiley[2,*], Stephen White[6,*], Julien Thézé[3,*], Diogo M Magnani[7], Karla Prieto[2], Daniel Reyes[2], Andrea Bingham[8], Lauren M Paul[5], Refugio Robles-Sikisaka[1], Glenn Oliveira[9], Darryl Pronty[6], Hayden C Metsky[10], Mary Lynn Baniecki[10], Kayla G Barnes[10], Bridget Chak[10], Catherine A Freije[10], Adrianne Gladden-Young[10], Andreas Gnirke[10], Cynthia Luo[10], Bronwyn MacInnis[10], Christian B Matranga[10], Daniel J Park[10], James Qu[10], Stephen F Schaffner[10], Christopher Tomkins-Tinch[10], Kendra L West[10], Sarah M Winnicki[10], Shirlee Wohl[10], Nathan L Yozwiak[10], Joshua Quick[11], Joseph R Fauver[12], Kamran Khan[13,14], Shannon E Brent[13], Robert C Reiner Jr.[15], Paola N Lichtenberger[16], Michael Ricciardi[7], Varian K Bailey[7], David I Watkins[7], Marshall R Cone[17], Edgar W Kopp IV[17], Kelly N Hogan[17], Andrew C Cannons[17], Reynald Jean[18], Robert F Garry[19], Nicholas J Loman[11], Nuno R Faria[3], Mario C Porcelli[20], Chalmers Vasquez[20], Elyse R Nagle[2], Derek AT Cummings[21], Danielle Stanek[8], Andrew Rambaut[22,23], Mariano Sanchez-Lockhart[2], Pardis C Sabeti[10,#], Leah D Gillis[6,#], Scott F Michael[5,#], Trevor Bedford[4,#], Oliver G Pybus[3,#], Sharon Isern[5,#], Gustavo Palacios[2,#,$], Kristian G Andersen[1,9,24,#,$]

[1]Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, CA 92037, USA. [2]Center for Genome Sciences, U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, MD 21702, USA. [3]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. [4]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. [5]Department of Biological Sciences, College of Arts and Sciences, Florida Gulf Coast University, Fort Myers, FL 33965, USA. [6]Bureau of Public Health Laboratories, Division of Disease Control and Health Protection, Florida Department of Health, Miami, FL 33125, USA. [7]Department of Pathology, University of Miami Miller School of Medicine, Miami, FL 33136, USA. [8]Bureau of Epidemiology, Division of Disease Control and Health Protection, Florida Department of Health, Tallahassee, FL 32399, USA. [9]Scripps Translational Science Institute, La Jolla, CA 92037, USA. [10]The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [11]Institute of Microbiology and Infection, University of Birmingham, Birmingham B15 2TT, UK. [12]Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO 80523, USA. [13]Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, ON M5B 1T8, Canada. [14]Division of Infectious Diseases, Department of Medicine, University of Toronto, Toronto, ON M5B 1T8, Canada. [15]Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA 98121, USA. [16]Division of Infectious Diseases, University of Miami Miller School of Medicine, Miami, FL 33155, USA. [17]Bureau of Public Health Laboratories, Division of Disease Control and Health Protection, Florida Department of Health, Tampa, FL 33612, USA. [18]Florida Department of Health in Miami-Dade County, Miami, FL 33125, USA. [19]Department of Microbiology and Immunology, Tulane University School of Medicine, New Orleans, LA 70112, USA. [20]Miami-Dade County Mosquito Control, Miami, FL 33178 USA. [21]Department of Biology and Emerging Pathogens Institute, University of Florida, Gainesville, FL 32610, USA. [22]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, UK. [23]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA. [24]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. [*] = co-first; [#] = co-senior; [$] = co-corresponding; K.G.A (andersen@scripps.edu); G.P. (gustavo.f.palacios.ctr@mail.mil).

Zika virus (ZIKV) is causing an unprecedented epidemic linked to severe congenital syndromes[1,2]. In July 2016, mosquito-borne ZIKV transmission was first reported in the continental United States and since then, hundreds of locally-acquired infections have been reported in Florida[3]. To gain insights into the timing, source, and likely route(s) of introduction of ZIKV into the continental United States, we tracked the virus from its first detection in Miami, Florida by direct sequencing of ZIKV genomes from infected patients and *Aedes aegypti* mosquitoes. We show that at least four distinct ZIKV introductions contributed to the outbreak in Florida and that local transmission likely started in the spring of 2016 - several months before its initial detection. By analyzing surveillance and genetic data, we discovered that ZIKV moved among transmission zones in Miami. Our analyses show that most introductions are phylogenetically linked to the Caribbean, a finding corroborated by the high incidence rates and traffic volumes from the region into the Miami area. By comparing mosquito abundance and travel flows, we describe the areas of southern Florida that are especially vulnerable to ZIKV introductions. Our study provides a deeper understanding of how ZIKV initiates and sustains transmission in new regions.

ZIKV transmission in the Americas was first reported in Brazil in May 2015[4], though the virus was likely introduced 1-2 years prior to its detection[5]. By January 2016, the ZIKV epidemic had spread to several South and Central American countries and most islands in the Caribbean[6]. Like dengue virus (DENV) and chikungunya virus (CHIKV), ZIKV is vectored primarily by *Aedes* mosquitoes[7–10]. The establishment of the peridomestic species *Ae. aegypti* in the Americas[11] has facilitated DENV, CHIKV, and now likely ZIKV to become endemic in this region[12]. In the continental United States, transient outbreaks of DENV and CHIKV have been reported in regions of Texas and Florida[13–19] with abundant seasonal *Ae. aegypti* populations[11,20].

In July 2016, the first locally-acquired ZIKV cases in the continental United States were reported in Florida[3]. The outbreak, which generated 256 confirmed ZIKV infections in 2016, was likely initiated by travel-associated infections in Florida (Fig. 1a). While transmission was confirmed across four counties in Florida (Fig. 1b), the ZIKV outbreak was most intense in Miami-Dade County (241 infections). Although the location of exposure could not be determined in all cases, at least 114 (47%) of the infections were likely acquired in one of three distinct transmission zones within Miami: Wynwood, Miami Beach, and Little River (Fig. 1c-d).
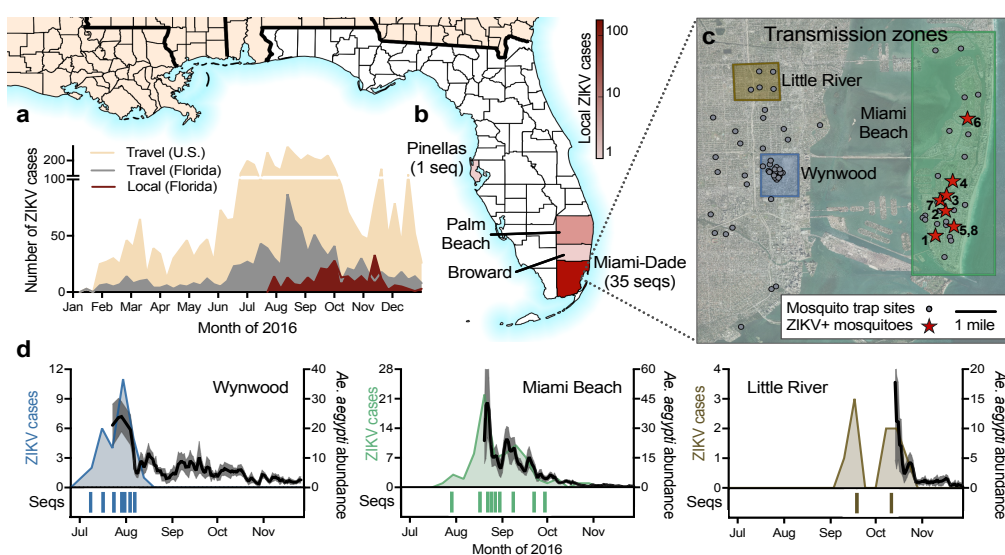

Figure 1. Zika virus outbreak in Florida. (**a**) Weekly counts of confirmed travel-associated and locally-acquired ZIKV cases in 2016. (**b**) Four counties reported locally-acquired ZIKV cases in 2016: Miami-Dade (241), Broward (5), Palm Beach (8), Pinellas (1), and unknown origin (1). (**c**) The locations of mosquito traps and collected *Ae. aegypti* mosquitoes found to contain ZIKV RNA (ZIKV+) in relation to the transmission zones within Miami. (**d**) Temporal distribution of weekly ZIKV cases (left y-axis), sequenced cases (bottom), and *Ae. aegypti* abundance per trap night (right y-axis) associated with the three described transmission zones. ZIKV cases and sequences are plotted in relation to symptom onset dates (n=18). Sequenced cases without onset dates or that occurred outside of the transmission zones are not shown (n=10). Human cases and *Ae. aegypti* abundance per week were positively correlated (Spearman r = 0.61, Supplementary Data Fig. 1b).

Using mosquito surveillance data, we determined the extent of mosquito-borne ZIKV transmission in Miami. Of the 24,351 mosquitoes collected throughout Miami-Dade County from June to November 2016, 99.8% were *Ae. aegypti* and 8 pools of *Ae. aegypti* collected in Miami Beach tested positive for ZIKV RNA (Fig. 1c, Supplementary Fig. 1). From these pools, we estimated that ~1 out of 1,600 *Ae. aegypti* mosquitoes were infected (0.061%, 95% CI: 0.028-0.115%, Supplementary Fig. 1a). This is similar to reported *Ae. aegypti* infection rates during DENV and CHIKV outbreaks[21]. Even though ZIKV-infected mosquitoes were not detected outside Miami Beach (Fig. 1c), we found that the number of human ZIKV cases correlated strongly with *Ae. aegypti* abundance within each of the defined transmission zones (Spearman r = 0.61, Fig. 1d, Supplementary Fig. 1b). This suggests that *Ae. aegypti* mosquitoes were the primary mode of transmission throughout Miami-Dade County and that changes to vector abundance directly impacted human infection rates. The application of insecticides[3], which we found suppressed local mosquito populations during the periods of intensive usage (Supplementary Fig. 1c), therefore likely contributed to the clearance of ZIKV within these zones.

To gain molecular insights into the Florida outbreak, we sequenced 39 ZIKV genomes directly from clinical and mosquito samples without cell culture enrichment[22] (Supplementary File 1). Our Florida ZIKV dataset (36 genomes) included 29 genomes from patients with locally-acquired infections (Fig. 1d) and 7 from *Ae. aegypti* mosquito pools (Fig. 1c). We also sequenced 3 ZIKV genomes from travel-associated cases diagnosed in Florida. Our dataset included cases from all three transmission zones in Miami (Fig. 1d) and represented ~11% of all confirmed locally-acquired cases

in Florida. We made all sequence data openly available at NCBI (BioProjects PRJNA342539, PRJNA356429) and other online resources immediately after data generation.

To infer viral genealogies, we reconstructed phylogenetic trees from our 39 ZIKV genomes along with 65 published genomes from other affected regions (Fig. 2, Supplementary Fig. 2 and 3). We found that the Florida ZIKV genomes formed four distinct lineages (labeled F1-F4, Fig. 2a), three of which (F1-F3) belonged to the same major clade (labeled A, Fig. 2a). We only sampled a single human case each from the F3 and F4 lineages, consistent with limited transmission (Fig. 2a). The other two Florida lineages (F1-F2) comprised ZIKV genomes sequenced from multiple human and mosquito samples within Miami-Dade County (Fig. 2b).
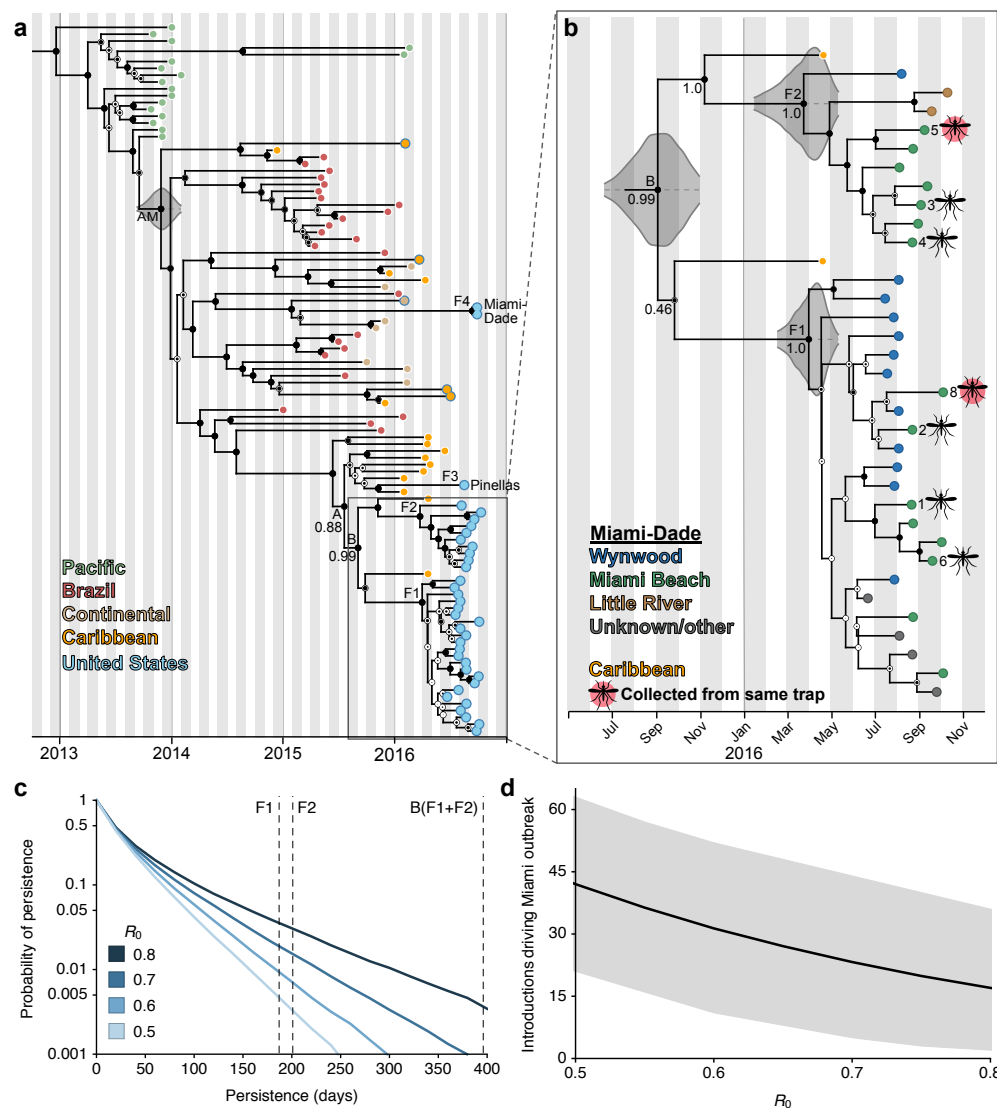
Figure 2. Multiple introductions of Zika virus into Florida. (**a**) Maximum clade credibility (MCC) tree of ZIKV genomes sequenced from outbreaks in the Pacific islands and the epidemic in the Americas. Tips are colored based on collection location. The five tips outlined in blue but filled with a different color indicate ZIKV cases in the United States associated with travel (fill color indicates the probable location of infection). Clade posterior probabilities are indicated with black circles inside white circles with posterior probability of 1 filling the entire circle black. The grey violin plot indicates the 95% highest posterior density (HPD) interval for the tMRCA for the epidemic in the Americas (AM)[5]. Lineage F4 contains two identical ZIKV genomes from the same patient. (**b**) A zoomed in version of the whole MCC tree showing the collection locations of Miami-Dade sequences and whether they were sequenced from mosquitoes (numbers correspond to trap locations in Fig. 1c). 95% HPD intervals are shown for the tMRCAs (**c**) The probability of ZIKV persistence after introduction for different R_0. Persistence is measured as the number of days from initial introduction of viral lineages until their extinction. Vertical dashed lines show the inferred mean persistence time for lineages F1, F2 and B based on their tMRCA. (**d**) Total number of introductions (mean with 95% CI) that contributed to the outbreak of 241 local cases in Miami-Dade County for different R_0.

Using time-structured phylogenies[23], we estimated that at least four separate introductions were responsible for the locally-acquired cases observed in our dataset. The phylogenetic placement of lineage F4 clearly indicates that it resulted from an independent introduction of a lineage genetically distinct from those in clade A (Fig. 2a). For the two well-supported nodes linking lineages F1 and F2 (labeled B, Fig. 2a) and F1-F3 (A, Fig. 2a), we estimated the time of the most recent common ancestor (tMRCA) to be during the summer of 2015. Our data displayed a strong clock signal (Supplementary Fig. 2b) and tMRCA estimates were robust across a range of different molecular clock and coalescent models (Supplementary Table 1). Thus while F1-F3 all belong to clade A, any less than three distinct introductions

leading to these lineages would have required undetected transmission of ZIKV in Florida for approximately one year (Fig. 2a).

To estimate the likelihood of a single ZIKV transmission chain persisting for over a year, we modeled epidemic spread under different assumptions of the basic reproductive number (R_0). We calculated R_0 using the locally-acquired and travel-associated case counts in Miami-Dade County, along with the number of observed genetic lineages. Allowing for variability in the percentage of infectious travel-associated cases and for sampling bias across lineages, we estimated an R_0 for Miami-Dade between 0.5 and 0.8 (Supplementary Fig. 4). Even at the upper end of this range, the probability of a single transmission chain persisting for over a year is extremely low (~0.5%, Fig. 2c). This is especially true considering the low *Ae. aegypti* abundance during the winter months (Supplementary Fig. 1d). Given the low probability of long-term persistence, we expect that the ZIKV genomes we sequenced from Florida (F1-F4) were the result of at least four separate introductions. Differences in surveillance practices across areas, combined with the high number of travel-associated cases in Florida (Fig. 1a), however, likely mean that unsampled ZIKV introductions contributed to the outbreak. To estimate the total number of underlying ZIKV introductions, we modeled epidemic scenarios that resulted in 241 locally-acquired cases within Miami-Dade County, and found that with R_0 values of 0.5-0.8, we expect 17-42 (95% CI 3-63) separate introductions to have contributed to the Miami-Dade outbreak (Fig. 2d). The majority of these introductions would likely have generated a single secondary case that was undetected in our genetic sampling (Supplementary Fig. 4a).

The two main ZIKV lineages, F1 and F2, included the vast majority of genomes we sampled in Florida (92%, Fig. 2a). Assuming they represent two independent introductions, we estimated when each of these lineages arrived in Florida. The probability densities for the tMRCAs of both F1 and F2 were centered around March-April, 2016 (Fig. 2b, 95% highest posterior density [HPD]: January-May, 2016). The estimated timing for these introductions corresponds with suitable *Ae. aegypti* populations in Miami-Dade County[24] (Supplementary Fig. 1d) and suggests that ZIKV transmission in Florida could have started at least two months prior to its first detection in July 2016 (Fig. 1a). The dates of the introductions could be more recent if multiple F1 or F2 lineage viruses arrived into Florida independently. However, more than 2 introductions from each lineage would be necessary to substantially change our estimates for the timing of the earliest introduction of ZIKV into Florida.

To better understand the transmission dynamics within Miami, we analyzed our genomic data together with case investigation data from the Florida Department of Health (DOH, Supplementary Table 1). While spatially distinct, the three ZIKV transmission zones occurred within ~3 miles of each other (Fig. 1c) and we found that the ZIKV infections associated with each zone overlapped temporally (Fig. 1d). Our 22 ZIKV genomes with zone assignments all belonged to lineages F1 and F2, but neither of these lineages were confined to a single transmission zone (Fig. 2b). In fact, we detected both F1 and F2 lineage viruses from *Ae. aegypti* collected from the same trap 26 days apart (mosquitoes 5 and 8, Fig. 2b). The detection of genetically similar viruses and the temporal overlap of ZIKV infections suggest that ZIKV moved among areas of Miami.

Given that ZIKV transmission is unlikely to persist through the winter in Florida[25] (Fig. 2c, Supplementary Fig. 1d), determining the sources and routes of ZIKV introductions could help mitigate future outbreaks. We found that lineages F1-F3 clustered with ZIKV genomes sequenced from the Dominican Republic and Guadeloupe (Fig. 2, Supplementary Fig. 2 and 3). In contrast, the F4 lineage clustered with genomes from Central America (Fig. 2, Supplementary Fig. 2 and 3). These findings suggest that while ZIKV outbreaks occurred throughout the Americas, the Caribbean islands appear to have been the main source of ZIKV that resulted in local transmission in Florida. Because of the severe undersampling of ZIKV genomes from across the Americas, however, we cannot rule out other source areas or panmixis of the ZIKV population in the region. Similarly, even though we found that the Florida ZIKV genomes clustered together with sequences from the Dominican Republic, our results do not prove that ZIKV entered Florida from this country.

We investigated ZIKV infection rates and travel patterns to corroborate our phylogenetic evidence for Caribbean introductions into Florida. We found that the Caribbean islands bore the highest ZIKV incidence rates around the likely time of the main introductions into Florida (Fig. 2b), despite Brazil and Colombia reporting the highest absolute number

of ZIKV cases (January to June, 2016, Fig. 3a, Supplementary Fig. 5, Supplementary File 2). During the same time period, ~3 million travelers arriving from the Caribbean accounted for 54% of the total traffic into Miami, with the vast majority (~2.4 million) arriving via cruise ships (Fig. 3b, Supplementary Fig. 6, Supplementary File 2). Combining the infection rates with travel capacities, we estimated that ~60-70% of ZIKV infected travelers entering Miami arrived from the Caribbean (Fig. 3c and Supplementary Fig. 7a). In support of these estimates, we found that the observed number of travel-associated ZIKV cases correlated strongly with the expected number of importations from the Caribbean (Spearman r = 0.8, Fig. 3d, Supplementary Fig. 7b). Also, 67% of the travel-associated infections diagnosed in Florida reported recent travel to the Caribbean (Fig. 3e); however, the mode of travel for these cases is unknown. Taken together, these findings suggest that a high incidence of ZIKV in the Caribbean, combined with frequent travel from the region, could have played a key role in the establishment of ZIKV transmission in Florida in the spring of 2016. These findings, however, do not indicate that cruise ships themselves are risk factors for human ZIKV infection, but only that they served as a major mode of transportation from areas with active transmission. In addition, the probability of ZIKV exposure may vary among individuals depending on their purpose of travel and therefore, we cannot determine the specific contribution of ZIKV-infected travelers arriving via airlines or cruise ships.
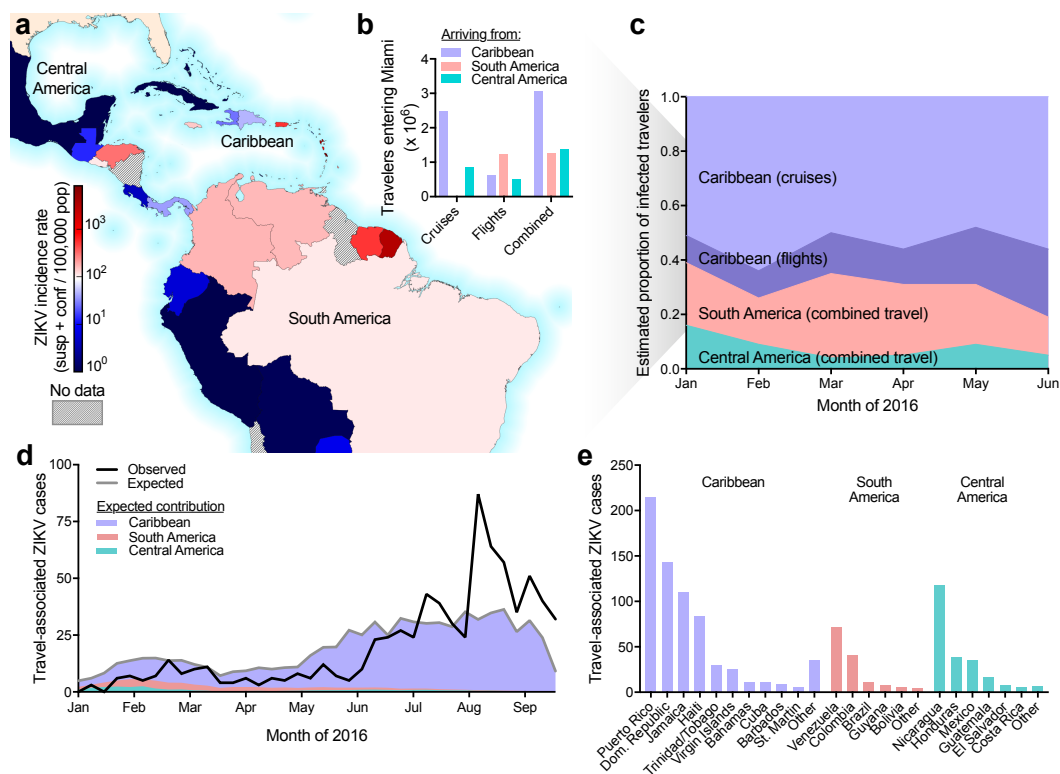
Figure 3. Frequent opportunities for Zika virus introductions into Miami from the Caribbean. (**a**) Reported ZIKV cases per country/territory from January to June, 2016 normalized by total population. (**b**) The number of estimated travelers entering Miami during January to June, 2016 by method of travel. (**c**) The number of travelers and the reported ZIKV incidence rate for the country/territory of origin were used to estimate the proportion of infected travelers coming from each region with ZIKV in the Americas. (**d**) The observed number of weekly travel-associated ZIKV cases in Florida were plotted with the expected number of ZIKV-infected travelers (as estimated in panel c) coming from all of the Americas (grey line) and the regional contributions (colored areas). (**e**) The countries visited by the 1,016 travel-associated ZIKV cases diagnosed in Florida.

The vast majority of the Florida ZIKV outbreak occurred in Miami-Dade County (94% of reported local cases in Florida, Fig. 1b). To determine if there is a higher potential for ZIKV outbreaks in the Miami area, we analyzed incoming passenger traffic by air and sea from regions with ZIKV transmission and *Ae. aegypti* abundance. We found that Miami and nearby Fort Lauderdale received ~72% of the traffic entering Florida (Fig. 4) and Miami received more air and sea traffic than any other city in the continental United States in 2016 (Supplementary Fig. 8). During January to April 2016, we estimated that *Ae. aegypti* abundance was highest in southern Florida[20] (Fig. 4, Supplementary Fig. 1d, Supplementary Fig. 8). By June, most of Florida and several cities across the South, such as New Orleans, Atlanta, and Charleston, were expected to support high seasonal *Ae. aegypti* populations[20] (Supplementary Fig. 8); however, none of these cities have reported local *Ae. aegypti*-borne transmission of any virus in at least 60 years[17]. In fact, despite the thousands of opportunities for introductions across the country (Fig. 1a), the only region outside of Florida that has reported mosquito-borne ZIKV transmission is southern Texas[27]. Southern Texas is also the only other region to recently

report DENV outbreaks[17–19]. Therefore, the combination of international travelers, mosquito ecology, and human population density near Miami likely make it one of the few places in the continental United States most at risk for *Ae. aegypti*-borne virus outbreaks[20,24,26].
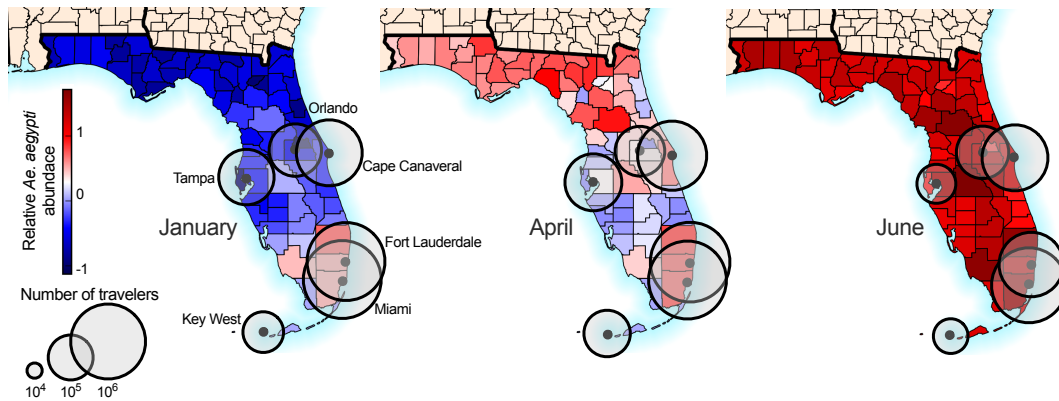
Figure 4. Southern Florida has a high potential for *Aedes aegypti*-borne virus outbreaks. The number of travelers per month (circles) entering Florida cities via flights and cruise ships were plotted with estimated relative *Ae. aegypti* abundance. Only cities receiving >10,000 passengers per month are shown. Relative *Ae. aegypti* abundance for every month is shown in Supplementary Fig. 1d.

The extent of ZIKV transmission in Florida was unprecedented, with more reported ZIKV cases in 2016 (256) than DENV cases since 2009 (136)[13]. Given that the majority of ZIKV infections are asymptomatic[2,28], the true number of ZIKV cases was likely much higher. Despite this, we estimated that the average R_0 was less than 1 and therefore multiple introductions were necessary to give rise to the magnitude of the observed outbreak[25]. The high volume of airline and cruise ship traffic entering Florida from ZIKV-affected regions, especially the Caribbean, likely provided a substantial supply of ZIKV-infected individuals[29,30]. Because Florida is unlikely to sustain long-term ZIKV transmission[25], the potential for future ZIKV outbreaks in this region is highly dependent upon activity elsewhere. Therefore, we expect that outbreaks in Florida will cycle with the ZIKV transmission dynamics in the Americas[12].

Acknowledgements

Author Contributions

Methods

Ethical statement

This study has been evaluated and determined to be exempt by the Institutional Review Boards (IRB) at The Scripps Research Institute (TSRI) and the USAMRIID Office of Human Use and Ethics. The study was reviewed by the Florida DOH Human Research Protection Program and determined not to require IRB review.

Florida Zika virus case data

Weekly reports of international travel-associated and locally-acquired ZIKV infections diagnosed in Florida were obtained from the Florida DOH mosquito-borne disease surveillance system[13]. Dates of symptom onset from the Miami transmission zones (Wynwood, Miami Beach, and Little River) determined by the Florida DOH investigation process were obtained from the ZIKV resource website[31] and daily updates[32]. International travel-associated ZIKV case counts in the United States (outside of Florida) were obtained from the CDC[33]. The local and travel-associated ZIKV case numbers for Florida were obtained from the Florida DOH. The one local ZIKV infection diagnosed in Duval County was believed to have originated elsewhere in Florida. Therefore, this case is listed as "unknown origin" in Fig. 1b. In Fig. 3e, only the countries visited 5 or more times by ZIKV-infected travelers diagnosed in Florida are shown. Countries with 5 or less visits were aggregated into an "other" category by region (*i.e.*, Caribbean, South America, or Central America).

Clinical sample collection

Clinical samples from locally-acquired ZIKV infections were collected from June 22 to October 11, 2016. The Florida DOH identified persons with compatible illness and clinical samples were shipped to the Bureau of Public Health Laboratories for confirmation by qRT-PCR and antibody tests following interim guidelines[3,34–36]. Clinical specimens (whole blood, serum, saliva, or urine) submitted for analysis were refrigerated or frozen at $\leq$ -70°C until RNA was extracted. RNA was extracted using the RNAeasy kit (QIAGEN), MagMAX for Microarrays Total RNA Isolation Kit (Ambion), or MagNA Pure LC 2.0 or 96 Systems (Roche Diagnostics). Purified RNA was eluted into 50-100 μL using the supplied elution buffers, immediately frozen at $\leq$ -70°C, and transported on dry ice. The Florida DOH also provided investigation data for these samples, including symptom onset dates and, when available, assignments to the zone where infection likely occurred (Supplementary File 1).

Mosquito collection and entomological data analysis

24,351 *Ae. aegypti* and *Ae. albopictus* mosquitoes (sorted into 2,596 pools) were collected throughout Miami-Dade County during June to November, 2016 using BG-Sentinel mosquito traps (Biogents AG). Up to 50 mosquitoes of the same species and sex were pooled per trap. The pooled mosquitoes were stored in RNAlater (Invitrogen), RNA was extracted using either the RNAeasy kit (QIAGEN) or MagMAX for Microarrays Total RNA Isolation Kit (Ambion), and ZIKV RNA was detected by qRT-PCR targeting the envelope protein coding region[36] or the Trioplex qRT-PCR kit[37]. ZIKV infection rates were calculated per 1,000 female *Ae. aegypti* mosquitoes using the bias-corrected maximum likelihood estimate (MLE)[38]. Days of insecticide usage by the Miami-Dade Mosquito Control were inferred from the zone-specific ZIKV activities timelines published by the Florida DOH[31].

Relative monthly *Ae. aegypti* abundance

For the purpose of this study we used *Ae. aegypti* suitability maps from Kraemer *et al.*[11] and derived monthly estimates based on the statistical relationships between mosquito presence and environmental correlates[41]. Following Hwang *et al.*[42] we used a simple mathematical formula to transform the probability of detection maps into mosquito abundance maps. In order to do so, we assumed P (Y=1) where Y is a binary variable (presence/absence). Using a Poisson distribution X() to govern the abundance of mosquitoes, the probability of not observing any mosquitoes can be related to the probability of absence as: P(X=0)=P(Y=0). We used the following transformation to generate abundance (λ) estimates per county in Florida:

$$e^{-\lambda} = P\left(Y = 0\right)$$
$$\lambda = -\log\left(P\left(Y = 0\right)\right)$$
$$\lambda = -\log\left(1 - P\left(Y = 1\right)\right)$$

Zika virus quantification

ZIKV genome equivalents (GE) were quantified by qRT-PCR. At TSRI, ZIKV qRT-PCR was performed as follows: ZIKV RNA standards were transcribed from the ZIKV NS5 region (8651-9498 nt) using the T7 forward primer (5' - TAA TAC GAC TCA CTA TAG GGA GA TCA GGC TCC TGT CAA AAC CC - 3'), reverse primer (5' - AGT GAC AAC TTG TCC GCT CC - 3'), and the T7 Megascript kit (Ambion). For qRT-PCR, primers and a probe targeting the NS5 region (9014-9123 nt) were designed using the ZIKV isolate PRVABC59 (GenBank: KU501215): forward primer (5'- AGT GCC AGA GCT GTG TGT AC - 3'), reverse primer (5' - TCT AGC CCC TAG CCA CAT GT - 3'), and FAM-fluorescent probe (5' - GGC AGC CGC GCC ATC TGG T - 3'). We have the best probes. The qRT-PCR assays were performed in 25 μl reactions using the iScript One-step RT-PCR Kit for probes (Bio-Rad Laboratories Inc.) and 2 μl of sample RNA. Amplification was performed at 50°C for 20 min, 95°C for 3 min, and 40 cycles of 95°C for 10 s and 57°C for 10 s. Fluorescence was read at the end of the 57°C annealing-extension step. 10-fold dilutions of the ZIKV RNA transcripts (2 μl/reaction) were used to create a standard curve for quantification of ZIKV GE/μl of RNA.

ZIKV GE were quantified at USAMRIID using the University of Bonn ZIKV envelope protein (Bonn E) qRT-PCR assay[43]. RNA standards were transcribed using an amplicon generated from a ZIKV plasmid containing T7 promoter at the start of the 5' untranslated region (UTR). The plasmid was designed using the ZIKV isolate BeH819015 (GenBank: KU365778.1) and the amplicon included nts 1-4348, which covers the 5' UTR, C, prM, M, E, NS1, and NS2 regions. The qRT-PCR assays were performed in 25 μl reactions using the SuperScript III platinum One-step qRT-PCR Kit (ThermoFisher) and 2 μl of sample RNA was used. Amplification was performed following conditions as previously described[43]. 10-fold dilutions of the ZIKV RNA transcripts (5 μl/reaction) were used to create a standard curve for quantification of ZIKV GE/μl of RNA.

Amplicon-based Zika virus sequencing

ZIKV sequencing at TSRI was performed using an amplicon-based approach using the ZikaAsian V1 scheme, as described[22]. Briefly, cDNA was reverse transcribed from 5 μl of RNA using SuperScript IV (Invitrogen). ZIKV cDNA (2.5 μl/reaction) was amplified in 35 × 400 bp fragments from two multiplexed PCR reactions using Q5 DNA High-fidelity Polymerase (New England Biolabs). The amplified ZIKV cDNA fragments (50 ng) were prepared for sequencing using the Kapa Hyper prep kit (Kapa Biosystems) and SureSelect XT2 indexes (Agilent). Agencourt AMPure XP beads (Beckman Coulter) were used for all purification steps. Paired-end 251 nt reads were generated on the MiSeq using the V2 500 cycle or V3 600 cycle kits (Illumina).

Trimmomatic was used to remove primer sequences (first 22 nt from the 5' end of the reads, which is the maximum length of the primers used for the multiplexed PCR) and bases at both ends with Phred quality score < 20[44]. The reads were then aligned to the complete genome of a ZIKV isolate from the Dominican Republic, 2016 (GenBank: KU853012) using Novoalign v3.04.04 (www.novocraft.com). Samtools was used to sort the aligned BAM files and to generate alignment statistics[45]. Snakemake was used as the workflow management system[46]. The code and reference indexes for the pipeline can be found at https://github.com/andersen-lab/zika-pipeline. ZIKV-aligned reads were visually inspected using Geneious v9.1.5[47] before generating consensus sequences. A minimum of 3× read-depth coverage, in support of the consensus, was required to make a base call.

The consensus ZIKV sequences from FL01M and FL03M generated by sequencing 35 × 400 bp amplicons on the MiSeq were validated using the following approaches: 1) sequencing the 35 × 400 bp amplicons on the Ion S5 platform (ThermoFisher), 2) sequencing amplicons generated using an Ion AmpliSeq® (ThermoFisher) panel customly targeted towards ZIKV on the Ion S5 platform, and 3) sequencing 5 × 2,150-2,400 bp ZIKV amplicons on the MiSeq. For Ion library preparation, cDNA was synthesized using the SuperScript VILO kit (ThermoFisher). ThermoFisher designed 875 custom ZIKV primers to produce 75 amplicons of ~200 bp in two PCR reactions for use with their Ion AmpliSeq Library Kit 2.0. The reagent FuPa was used to digest the modified primer sequences after amplification. The DNA

templates were loaded onto Ion 520 chips using the Ion Chef and sequenced on the Ion S5 with the 200 bp output (ThermoFisher). The 35 × 400 bp amplicons generated for the MiSeq as described above were introduced into the Ion workflow using the Ion AmpliSeq Library Kit 2.0, but without fragmentation. Primers to amplify 2,150-2,400 bp ZIKV fragments (Supplementary File 3) were kindly provided by Shelby O'Connor, Dawn Dudly, Dave O'Connor, and Dane Gellerup (AIDS Vaccine Research Laboratory, University of Wisconsin, Madison). Each fragment was amplified individually by PCR using the cDNA generated above, Q5 DNA High-fidelity Polymerase, and the following thermocycle conditions: 55 ℃ for 30 m, 94 ℃ for 2 m, 35 cycles of 94 ℃ for 15 s, 56 ℃ for 30 s, and 68 ℃ for 3.5 m, 68 ℃ for 10 m, and held at 4 ℃ until use. Each PCR product was purified using Agencourt AMPure XP beads, sheared to 300 to 400 nt fragments using the Covaris S2 sonicator, indexed and prepared for sequencing as described above, and sequenced using the MiSeq V2 500 cycle kit (paired-end 251 nt reads). Compared to the consensus sequences generated using 35 × 400 bp amplicons on the MiSeq, there were no consensus-level mismatches in the coding sequence using any of the other three approaches (Supplementary Table 2). There were, however, some mismatches in the 5' and 3' UTRs (where the genomic RNA is heavily structured), likely a result of PCR bias and decreased coverage depth.

Enrichment-based Zika virus sequencing

ZIKV sequencing at USAMRIID was performed using a targeted enrichment approach. Sequencing libraries were prepared using the TruSeq RNA Access Library Prep kit (Illumina) with custom ZIKV probes. The set included 866 unique probes each of which was 80 nt in length (Supplementary File 3). The probes were designed to cover the entire ZIKV genome and to encompass the genetic diversity present on GenBank on January 14, 2016. In total, 26 ZIKV sequences were used during probe design (Supplementary File 3). Extracted RNA was fragmented at 94 °C for 0-60 s and each sample was enriched separately using a quarter of the reagents specified in the manufacturer's protocol. Samples were barcoded, pooled and sequenced using the MiSeq Reagent kit v3 (Illumina) on an Illumina MiSeq with a minimum of 2 × 151 bp reads. Dual indexing, with no overlapping indices, was used.

The random hexamer associated with read one and the Illumina adaptors were removed from the sequencing reads using Cutadapt v1.9.dev1[48], and low-quality reads/bases were filtered using Prinseq-lite v0.20.3[49]. Reads were aligned to a reference genome (GenBank: KX197192.1) using Bowtie2 v2.0.6[50], duplicates were removed with Picard (http://broadinstitute.github.io/picard), and a new consensus was generated using a combination of Samtools v0.1.18[45] and custom scripts (https://github.com/jtladner/Scripts/blob/master/reference-based_assembly/consensus_fasta.py). Only bases with Phred quality score ≥ 20 were utilized in consensus calling, and a minimum of 3× read-depth coverage, in support of the consensus, was required to make a call; positions lacking this depth of coverage were treated as missing (*i.e.* called as ''N'').

Phylogenetic analyses

All published and available complete ZIKV genomes of the Asian genotype from the Pacific and the Americas were retrieved from GenBank public database as of December 2016. Public sequences (n=65) were codon-aligned together with ZIKV genomes generated in this study (n=39) using MAFFT[51] and inspected manually . The multiple alignment contained 104 ZIKV sequences collected between 2013 and 2016, from the Pacific (American Samoa, French Polynesia, and Tonga), Brazil, other South and Central Americas (Guatemala, Mexico, Suriname, and Venezuela), the Caribbean (Dominican Republic, Guadeloupe, Haiti, Martinique, and Puerto Rico), and the United States (Supplementary File 1).

In order to determine the temporal signal of the sequence dataset, a maximum likelihood (ML) phylogeny was first reconstructed with PhyML[52] using the general time-reversible (GTR) nucleotide substitution model and gamma distributed rates amongst sites[54] (Supplementary File 1), which was identified as the best fitting model for ML inference by jModelTest2[55]. Then, a correlation between root-to-tip genetic divergence and date of sampling was conducted in TempEst[56].

Bayesian phylogenetic analyses were performed using BEAST v.1.8.4[23] to infer time-structured phylogenies. We used an SDR06 nucleotide substitution model[57] with a non-informative continuous time Markov chain reference prior (CTMC)[58] on the molecular clock rate. Replicate analyses using multiple combinations of molecular clock and coalescent demographic models were explored to select the best fitting model by marginal likelihood comparison using

path-sampling and stepping-stone estimation approaches[59–61] (Supplementary Table 3). The best fit model was a strict molecular clock along with a Bayesian skyline demographic model[62]. All the Bayesian analyses were run for 30 million Markov chain Monte Carlo steps, sampling parameters and trees every 3000 generations (BEAST XML file and MCC tree available in Supplementary File 1).

Expected number and distribution of local cases from Zika virus importations

We used branching process theory[63,64] to generate the offspring distribution (subsequent local cases) that is expected from a single introduction. The offspring distribution L is modelled with a negative binomial distribution with mean R_0 and over-dispersion parameter k. The total number of cases j that is caused by a single importation (including the index case) after an infinite time[65] has the following form:

$$L = \frac{\Gamma\left(kj + j - 1\right)}{\Gamma\left(kj\right)\Gamma\left(j + 1\right)} \frac{\left(\frac{R_0}{k}\right)^{j-1}}{\left(1 + \frac{R_0}{k}\right)^{kj+j-1}}$$

The parameter k represents the variation in the number of secondary cases generated by each case of ZIKV[63]. In the case of vector borne diseases, local heterogeneity is high due to a variety of factors such as mosquito population abundance, human to mosquito interaction, and control interventions[66,67]. Here, we assumed high heterogeneity (k=0.1) following previous estimates for vector borne diseases[64]. This distribution L is plotted in Supplementary Fig. 4a. For the following, we took a forward simulation approach, drawing random samples from this distribution. All estimates were based on 100,000 random simulations.

We used this formula to estimate the probability of observing 241 local cases in Miami-Dade County alongside 320 travel-associated cases. We approached this by sampling 320 introduction events from L and calculating the total number of local cases in the resulting outbreak (Supplementary Fig. 4b). We also calculated the likelihood of observing 241 local cases in the total outbreak (Supplementary Fig. 4c), finding that the MLE of R_0 lies between 0.35 and 0.55. As a sensitivity analysis, we additionally modelled introductions with the assumption that only 50% of travelers were infectious at time of arrival into Miami-Dade County, resulting in an MLE of R_0 of 0.45–0.8.

We further used this formula to address the probability of observing 3 distinct genetic clusters (F1, F2 and F3) representing 3 introduction events in a sample of 27 ZIKV genomes from Miami-Dade County. We approached this by sampling introduction events until we accumulated 241 local cases according to L, arriving at N introduction events with case counts ($j_1, j_2, \ldots j_N$). We then sampled 27 cases *without replacement* from ($j_1, j_2, \ldots j_N$) following a hypergeometric distribution and recorded the number of distinct clusters drawn in the sample. We found that higher values of R_0 resulted in fewer distinct clusters within the sample of 27 genomes (Supplementary Fig. 4d). We additionally calculated the likelihood of sampling 3 distinct genetic clusters in 27 genomes (Supplementary Fig. 4e), finding an MLE estimate of R_0 of 0.7–0.9. Additionally, as a sensitivity analysis we modelled a preferential sampling process in which larger clusters are more likely to be drawn from than smaller clusters. Here, we used a parameter α that enriches the hypergeometric distribution following ($j_1^{\alpha}, j_2^{\alpha}, \ldots j_N^{\alpha}$). In this case, we found an MLE estimate of R_0 of 0.5–0.9.

Using the overlap of estimates of R_0 from local case counts (0.35–0.8) and genetic clusters (0.5–0.9), we arrived at a 95% uncertainty range of R_0 of 0.5–0.8.

We additionally perform birth-death stochastic simulations assuming a serial interval with mean 20 days[12]. We record the number of stochastic simulations still persisting after a particular number of days for different values of R_0 (Fig. 2c).

Incidence and attack rates

Weekly suspected and confirmed ZIKV case counts from countries and territories within the Americas with local transmission (January 1 to September 18, 2016) were obtained from the Pan American Health Organization (PAHO)[68]. In most cases, the weekly case numbers per country were only reported in bar graphs, therefore we used WebPlotDigitizer v3.10 (http://arohatgi.info/WebPlotDigitizer) to estimate the numbers. Country and territory total

population sizes to calculate weekly and monthly ZIKV incidence rates were also obtained from PAHO[69]. Incidence rates calculated from countries and territories in the Americas during January to June, 2016 (based on the earliest introduction time estimates until the first known cases) were used as an estimate for infection likelihood to investigate sources of ZIKV introductions.

To account for potential reporting biases with incidence rates, we also used ZIKV attack rates (*i.e.*, proportion infected before epidemic burnout) to estimate infection likelihood. Attack rates were calculated using a susceptible–infected–recovered (SIR) transmission model derived from seroprevalence studies and environmental factors as described[70]. Using attack rates as an estimate of infection likelihood, we predict that ~60% of the infected travelers entering Miami came from the Caribbean (Supplementary 7b), which is in agreement with our methods using incidence rates of ~60-70% (Fig. 3c). A list of countries and territories used in these analyses can be found in Supplementary File 2.

Airline and cruise ship traffic

To investigate whether the transmission of ZIKV in Florida coincides with travel patterns from ZIKV endemic regions, we obtained the number of passengers arriving at airports in Florida via commercial air travel. We collated flight data from countries and territories in the Americas with local ZIKV transmission between January and June, 2016 (based on the earliest introduction time estimates until the first known cases, Supplementary File 2), arriving at all commercial airports in Florida. The data were obtained from the International Air Transportation Association, which collects data on an estimated 90% of all passenger trips worldwide. Nelson et al.[26] previously reported flight data from 33 countries with ZIKV transmission entering major United States airports during October 2014 through September 2015, which we used to assess the potential for ZIKV introductions outside of Florida.

Schedules for cruise ships visiting Miami, Port Canaveral, Port Everglades, Fort Lauderdale, Key West, Jacksonville (all in Florida), Houston, Galveston (both in Texas), Charleston (South Carolina) and New Orleans (Louisiana) ports in the year 2016 were collated from www.cruisett.com and confirmed by cross-referencing ship logs reported by Port of Miami and reported ship schedules from www.miamidade.gov/portmiami/. Scheduled cruise ship capacities were extracted from www.cruisemapper.com. Every country/territory with ZIKV transmission visited by a cruise ship 10 days (the approximate mean time to ZIKV clearance in human blood [*i.e.*, the infectious period])[71] prior to arrival was counted as contributing the ship's capacity worth of passengers to Miami to the month of arrival (Supplementary File 2).

Expected number of travelers infected with Zika virus

We estimated the expected number of travelers entering Miami who were infected with ZIKV (λ) by using the total travel capacity (C) and the likelihood of ZIKV infection (infections (I) per person (N)) from each country/territory (i):

$$\lambda = \sum_i C_i \frac{I_i}{N_i}$$

We summed the number of expected infected travelers from each country/territory with ZIKV transmission by region and travel method (flights or cruises). We reported the data as the relative proportion of infected travelers from each region (Fig. 3c, Supplementary Fig. 7a) and as the absolute number of infected travelers (Fig. 3d, Supplementary Fig. 7b, Supplementary File 2).

Supplementary Data

a Miami-Dade mosquito surveillence

Location	Start date	End date	Trap sites	Trap nights	*Ae. aegypti* collected	No. per trap night	Pools tested	ZIKV+ pools	MLE infection rate
Wynwood	Jul 23	Nov 25	29	824	4,972	6.0	738	0	0
Miami Beach	Aug 21	Nov 28	29	2,180	13,265	6.1	1,473	8	0.61 (0.28-1.15)
Little River	Oct 15	Nov 28	5	217	485	2.2	136	0	0
Other Miami	Jun 21	Sep 23	52	235	5,584	23.8	249	0	0
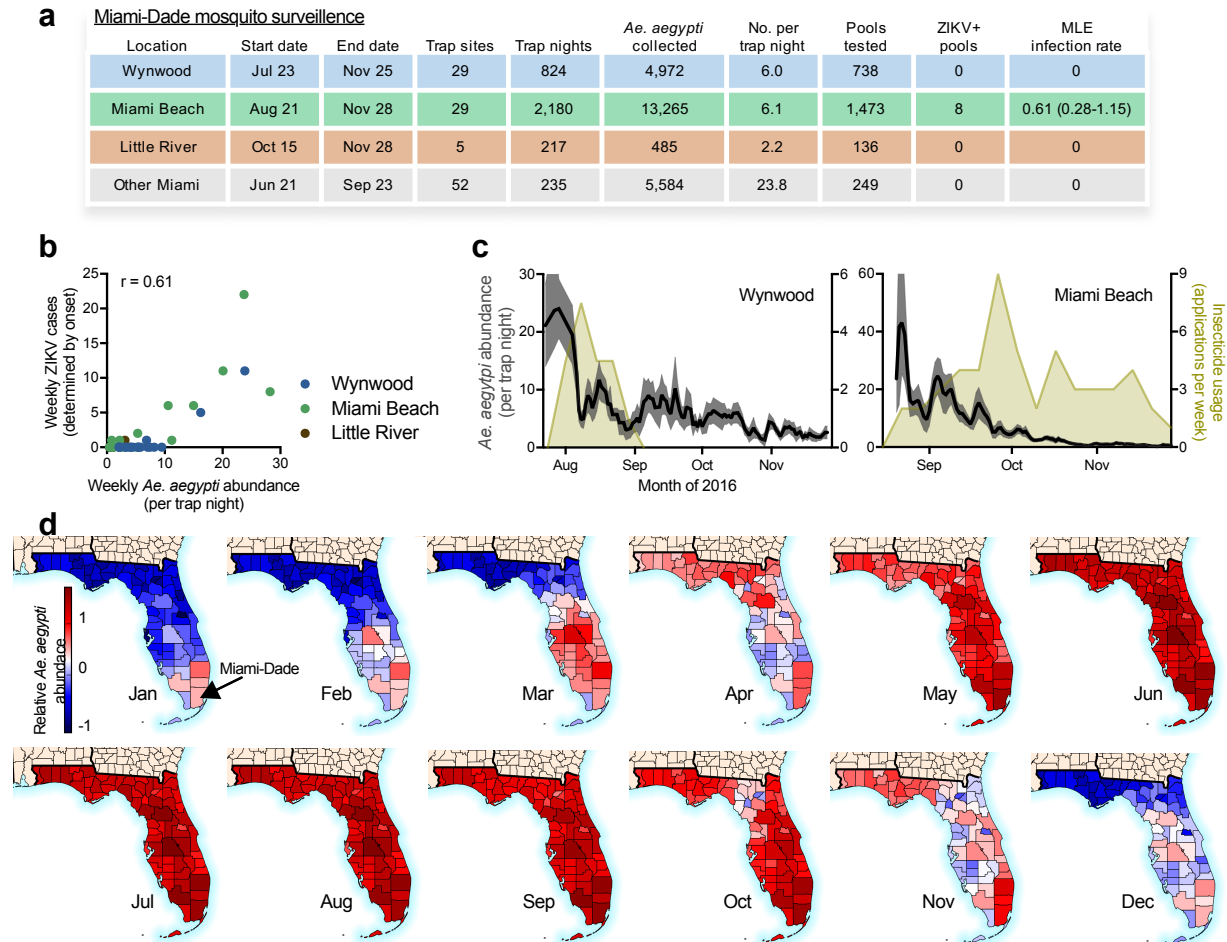
Supplementary Figure 1. Miami-Dade mosquito surveillance and relative *Aedes aegypti* abundance. (**a**) Mosquito surveillance data reported from June 21 to November 28, 2016 was used to evaluate the risk of ZIKV infection from mosquito-borne transmission in Miami. A total of 24,306 *Ae. aegypti* and 45 *Ae. albopictus* were collected. Trap nights are the total number of times each trap site was used and the trap locations are shown in Fig. 1d (some "Other Miami" trap sites are located outside of mapped region). Up to 50 mosquitoes of the same species and trap night were pooled together for ZIKV RNA testing. The infection rates were calculated using a maximum likelihood estimate (MLE). None of the *Ae. albopictus* pools contained ZIKV RNA. (**b**) The number of weekly ZIKV cases (based on symptoms onset) was correlated with mean *Ae. aegypti* abundance per trap night determined from the same week and zone (Spearman r = 0.61). This suggests that when the virus is present, mosquito abundance numbers alone could be used to target control efforts. (**c**) Insecticide usage, including truck and aerial adulticides and larvacides, by the Miami-Dade Mosquito Control in Wynwood (left) and Miami Beach (right) was overlaid with *Ae. aegypti* abundance per trap night to demonstrate that intense usage of insecticides may have helped to reduce local mosquito populations. (**d**) Relative *Ae. aegypti* abundance for each Florida county and month was estimated using a multivariate regression model, demonstrating spatial and temporal heterogeneity for the risk of ZIKV infection.
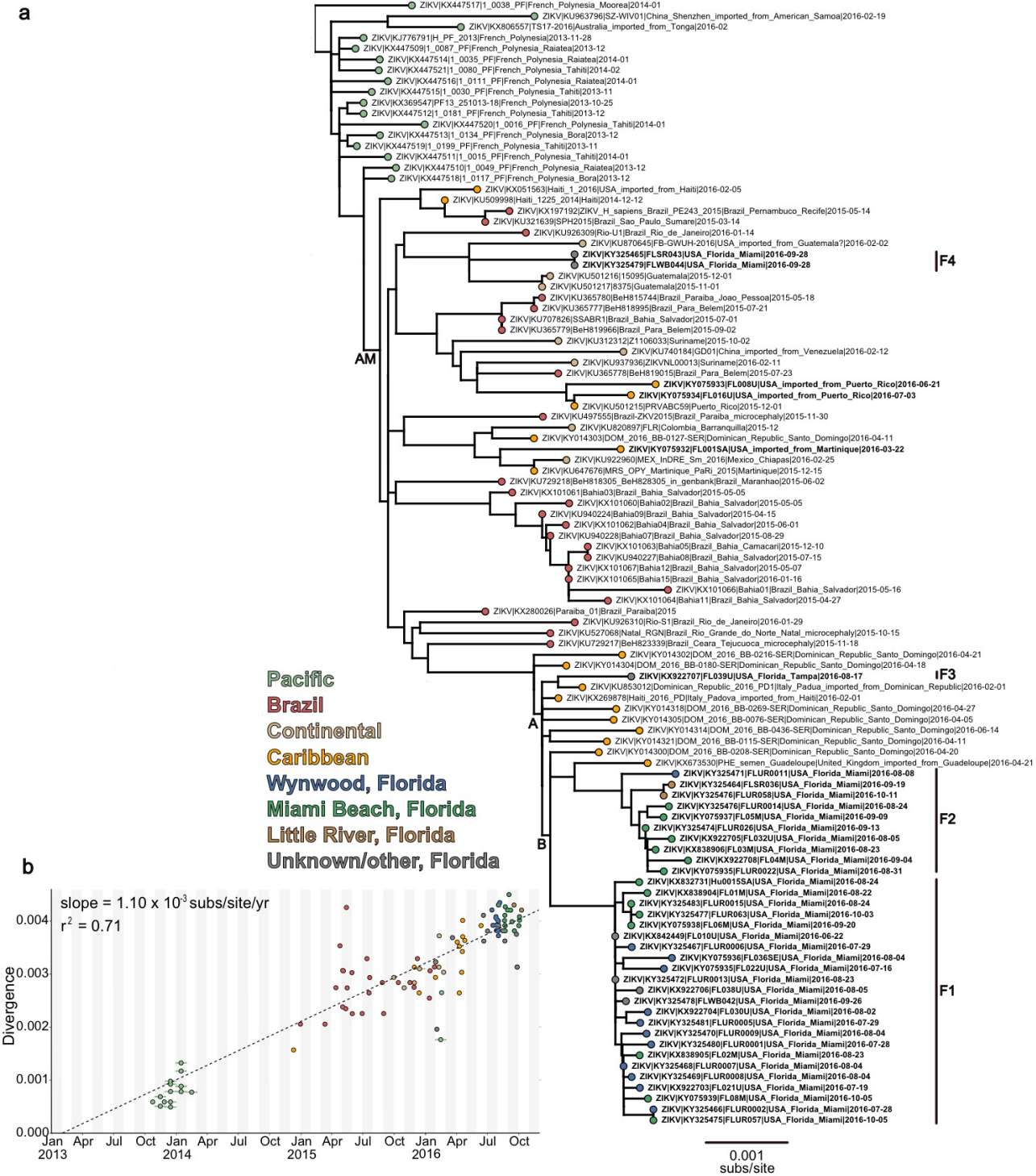
Supplementary Figure 2. Maximum likelihood tree and root-to-tip regression of Zika virus genomes from Pacific islands and the epidemic in Americas. (**a**) Maximum likelihood tree of publicly available ZIKV sequences and sequences generated in this study (n=104). tips are coloured by location, labels in bold indicate sequences generated in this study, Florida clusters F1-F4 are indicated by vertical lines to the right of the tree. (**b**) Linear regression of sample tip dates against divergence from root based on sequences with known collection dates estimates an evolutionary rate for the ZIKV phylogeny of 1.10×10^{-3} nucleotide substitutions/site/year (subs/site/yr). This is consistent with BEAST analyses using a strict molecular clock and a Bayesian skyline tree prior, the best-performing combination of clock and demographic model according to marginal likelihood estimates (Supplementary Table 3), which estimated an evolutionary rate of 1.12×10^{-3} (95% highest posterior density: $0.96 - 1.27 \times 10^{-3}$) subs/site/yr (Supplementary Table 1). These values are in agreement with previous estimates calculated based on ZIKV genomes from Brazil[5].

Supplementary Figure 3. Molecular clock dating of Zika virus clades. Maximum clade credibility (MCC) tree of ZIKV genomes collected from Pacific islands and the epidemic in Americas (n=104). Circles at the tips are colored based on origin location. Clade posterior probabilities are indicated with black circles inside white circles with posterior probability of 1 filling the entire circle black. The grey violin plot indicates the 95% highest posterior density (HPD) interval for the tMRCA of the American epidemic. We estimated that the tMRCA for the ongoing epidemic in the Americas occurred during October, 2013 (node AM, Extended Table 1, 95% HPD: August, 2013-January, 2014), which is consistent with previous analysis based on ZIKV genomes from Brazil[5].
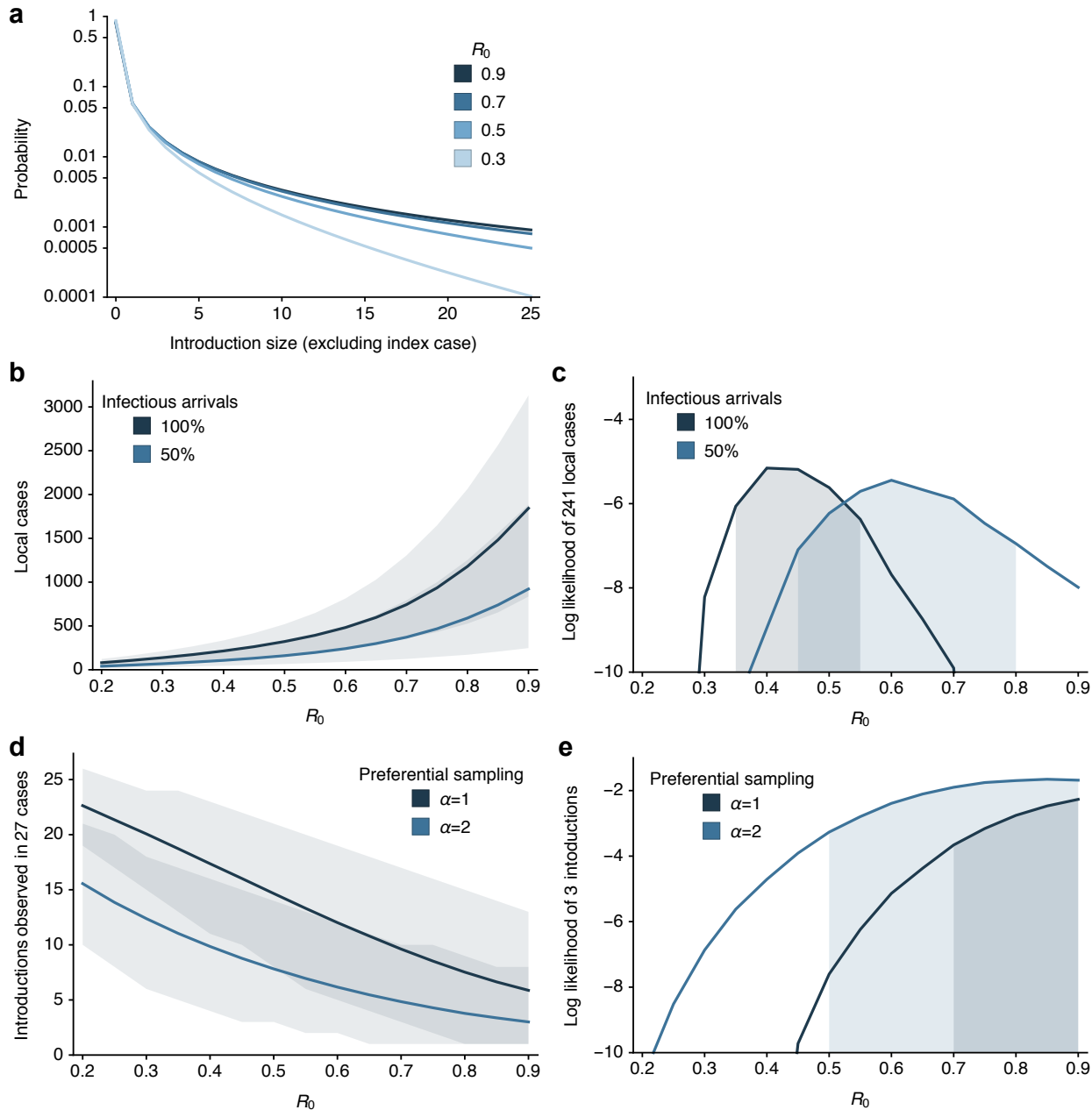
Supplementary Figure 4. Estimation of basic reproductive number and number of introductions. (**a**) Probability distribution of estimated total number of cases caused by a single introduction (excluding the index case) for different values of R_0. (**b**) Mean and 95% CI for total number of local cases caused by 320 introduction events (*i.e.*, travel-associated cases diagnosed in Miami-Dade County) for different values of R_0 and for different assumptions of proportion of infectious travelers. (**c**) Log likelihood of observing 241 local cases with 320 introduction events for different values of R_0 along with 95% maximum likelihood estimate (MLE) bounds on R_0. (**d**) Mean and 95% uncertainty interval for total number of distinct phylogenetic clusters observed in 27 sequenced ZIKV genomes for different values of R_0 and for different assumptions of sampling bias, from $\alpha=1$ (no sampling bias) to $\alpha=2$ (skewed toward preferentially sampling larger clusters). (**e**) Log likelihood of observing 3 clusters in 27 sequenced cases for different values of R_0 along with 95% MLE bounds on R_0.
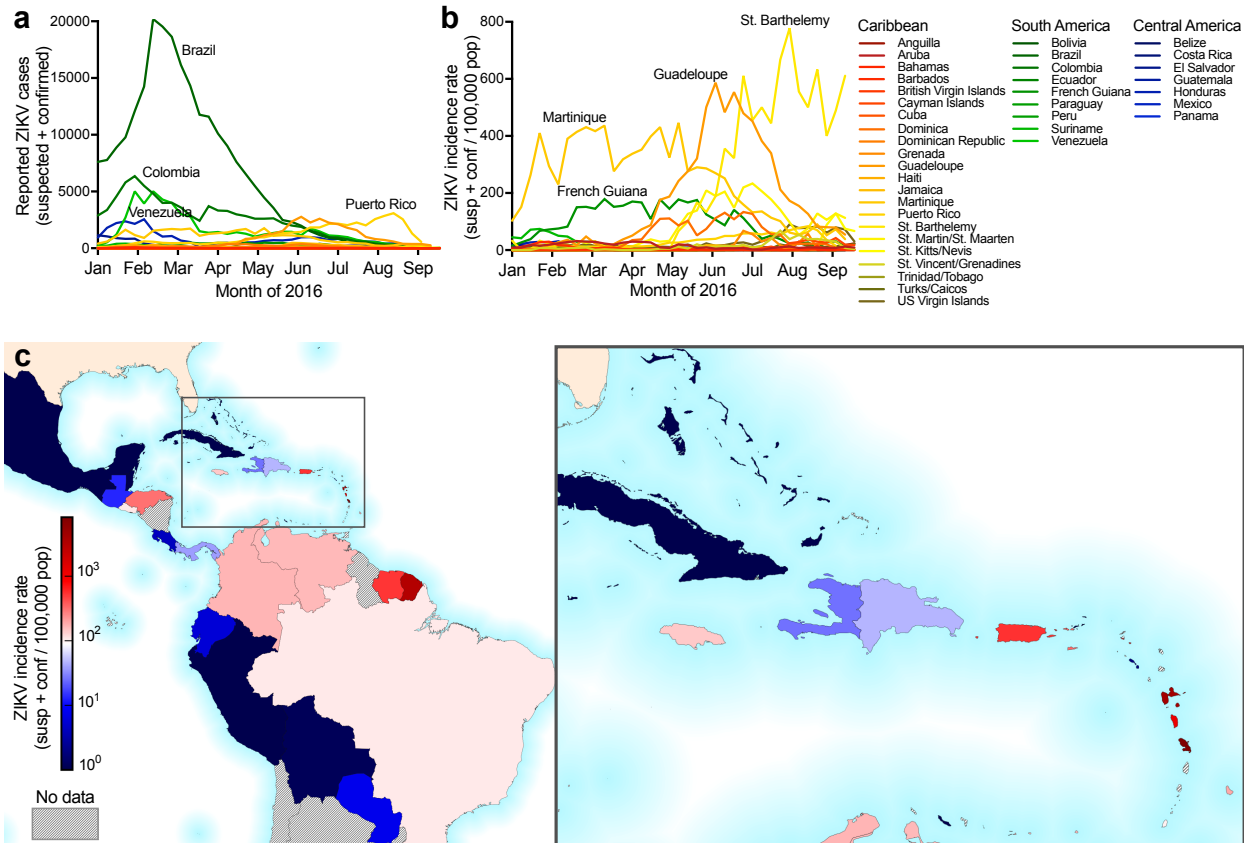
Supplementary Figure 5. Weekly reported Zika virus case numbers and incidence rates in the Americas. (**a**) ZIKV cases (suspected and confirmed) and (**b**) incidence rates (normalized per 100,000 population) are shown for each country or territory with available data per epidemiological week from January 1 to September 18, 2016. (**c**) Each country or territory with available data is colored by its reported ZIKV incidence rate from January to June, 2016 (the time frame for analysis of ZIKV introductions into Florida).
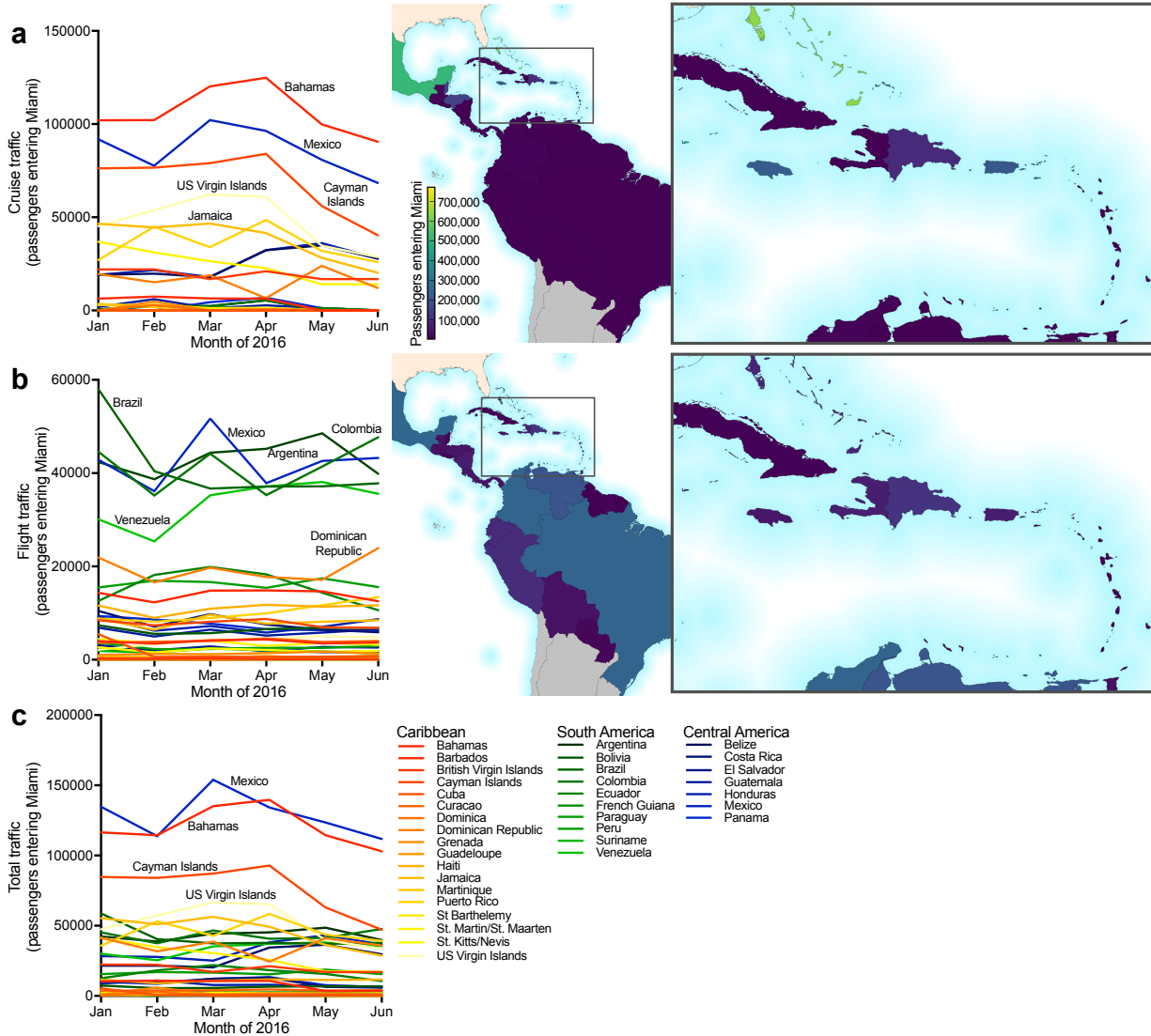
Supplementary Figure 6. Cruise and flight traffic entering Miami from regions with Zika virus transmission. The expected number of passengers entering Miami, by either (**a**) cruises or (**b**) flights, from each country or territory in the Americas with ZIKV transmission per month (left panel). The center map and inset show the cumulative numbers of travelers entering Miami during January to June, 2016 (the time frame for analysis of ZIKV introductions into Florida) from each country or territory per method of travel. (**c**) The total traffic (*i.e.* cruises and flights) is shown entering Miami per month.
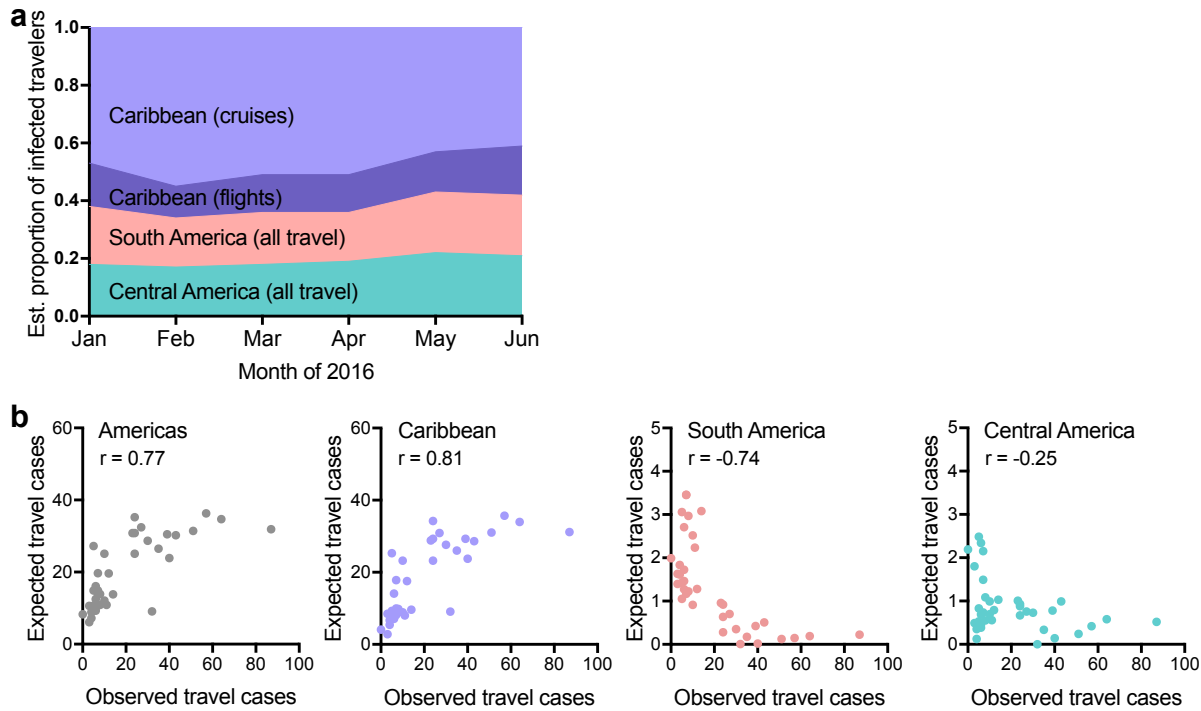
Supplementary Figure 7. Expected number of Zika virus infected travelers from the Caribbean is correlated with the total observed number of travel-associated infections. (**a**) In order to account for potential biases in ZIKV reporting accuracies, we also estimated the proportion of infected travelers using projected ZIKV attack rates[70] (*i.e.* predicted proportion of population infected before epidemic burnout). About 60% of the infected travelers are expected to have arrived from the Caribbean, similar to our results using incidence rates (Fig. 3c). (**b**) The expected number of travel-associated ZIKV cases were estimated by the number of travelers coming into Miami from each country/territory (travel capacity) and the in-country/territory infection likelihood (incidence rate per person) per week. The expected travel cases were summed from all of the Americas (left), Caribbean (left center), South America (right center), and Central America (right) and plotted with the observed travel-associated ZIKV cases. Numbers in each plot indicate Spearman correlation coefficients. Negative Spearman r coefficients indicated a negative correlation between the number of expected and observed travel cases.
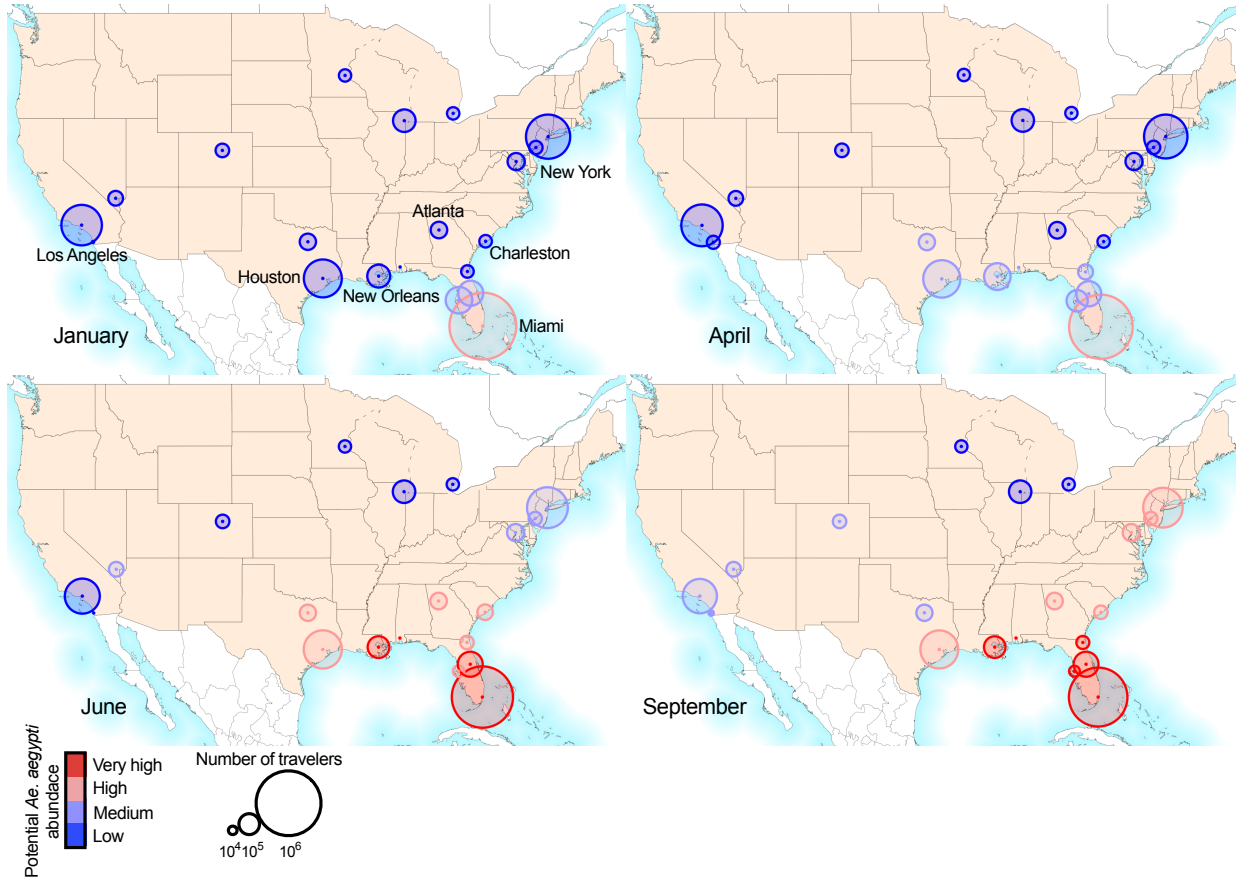
Supplementary Figure 8. Greater early season potential for Zika virus introductions into Miami. (**a**) The monthly cruise ship and airline[26] capacity from countries/territories with ZIKV transmission for the major United States travel hubs (shown as circle diameter) with monthly potential *Ae. aegypti* abundance (circle color), as previously estimated[20]. Cruise capacities from Houston and Galveston, Texas were combined.

Supplementary Table 1. Times of the most recent common ancestor and evolutionary rates

Model combination	Clade AM tMRCA			Clade A tMRCA			Clade B tMRCA		
	Mean	Lower 95% HPD	Lower 95% HPD	Mean	Lower 95% HPD	Lower 95% HPD	Mean	Lower 95% HPD	Lower 95% HPD
Strict, Constant	2013.88	2013.66	2014.09	2015.60	2015.44	2015.76	2015.71	2015.54	2015.87
Strict, Exponential	2013.90	2013.68	2014.08	2015.54	2015.35	2015.71	2015.65	2015.46	2015.82
Strict, Bayesian SkyGrid	2013.88	2013.65	2014.08	2015.59	2015.39	2015.75	2015.70	2015.49	2015.86
Strict, Bayesian Skyline	2013.92	2013.68	2014.12	2015.55	2015.35	2015.73	2015.69	2015.47	2015.85
UCLN, Constant	2013.96	2013.67	2014.23	2015.72	2015.49	2015.90	2015.83	2015.60	2016.03
UCLN, Exponential	2013.97	2013.73	2014.22	2015.65	2015.46	2015.82	2015.76	2015.56	2015.94
UCLN, Bayesian SkyGrid	2013.93	2013.65	2014.18	2015.69	2015.47	2015.87	2015.80	2015.59	2015.99
UCLN, Bayesian Skyline	2013.98	2013.71	2014.24	2015.61	2015.38	2015.85	2015.75	2015.49	2015.96

Model combination	Clade F1 tMRCA			Clade F2 tMRCA			Evolutionary Rate		
	Mean	Lower 95% HPD	Lower 95% HPD	Mean	Lower 95% HPD	Lower 95% HPD	Mean	Lower 95% HPD	Lower 95% HPD
Strict, Constant	2016.16	2016.03	2016.28	2016.17	2016.01	2016.32	1.09E-03	9.45E-04	1.25E-03
Strict, Exponential	2016.11	2015.95	2016.24	2016.13	2015.95	2016.28	1.07E-03	9.30E-04	1.22E-03
Strict, Bayesian SkyGrid	2016.16	2016.03	2016.28	2016.16	2015.99	2016.31	1.09E-03	9.30E-04	1.24E-03
Strict, Bayesian Skyline	2016.27	2016.14	2016.37	2016.24	2016.06	2016.38	1.12E-03	9.63E-04	1.27E-03
UCLN, Constant	2016.19	2016.06	2016.31	2016.22	2016.06	2016.37	1.24E-03	9.85E-04	1.48E-03
UCLN, Exponential	2016.15	2016.00	2016.28	2016.17	2016.01	2016.33	1.19E-03	9.80E-04	1.40E-03
UCLN, Bayesian SkyGrid	2016.18	2016.04	2016.31	2016.21	2016.03	2016.35	1.20E-03	9.98E-04	1.43E-03
UCLN, Bayesian Skyline	2016.29	2016.16	2016.39	2016.27	2016.09	2016.40	1.20E-03	1.01E-03	1.43E-03

HPD, highest posterior density. Dates listed as proportion of days elapsed with a year. Clades refer to Fig. 2a

Supplementary Table 2. Validation of sequencing results.

Sample	Amplicon method	NGS platform	Mismatches/nucleotides covered[a]		
			3' UTR	CDS	5' UTR
FL01M	35 × 400 bp	Ion S5	1/80	0/10272	7/252
	75 × ~200 bp[b]	Ion S5	2/75	0/10272	4/205
	5 × ~2,200 bp	MiSeq	0/80	0/10272	0/32
FL03M	35 × 400 bp	Ion S5	3/87	0/10272	20/252
	75 × ~200 bp[b]	Ion S5	4/78	0/10272	5/198
	5 × ~2,200 bp	MiSeq	0/82	0/10272	0/32

[a] Compared to the consensus genomes generated by sequencing 35 × 400 bp amplicons on the MiSeq.
[b] Amplicons produced using Ion AmpliSeq and 875 custom ZIKV primers.
NGS, next-generation sequencing; UTR, untranslated region; CDS, coding sequence.

Supplementary Table 3. Model selection to infer time-structured phylogenies.

Model combination	Path Sampling	Ranking	Stepping Stone	Ranking
Strict, Constant	-20659.122	3	-20663.265	3
Strict, Exponential	-20657.526	2	-20662.632	2
Strict, Bayesian SkyGrid	-20659.701	4	-20663.981	4
Strict, Bayesian Skyline	-20630.097	1	-20633.851	1
UCLN, Constant	-21749.149	7	-21754.349	7
UCLN, Exponential	-21737.820	6	-21741.818	6
UCLN, Bayesian SkyGrid	-21749.639	8	-21754.556	8
UCLN, Bayesian Skyline	-21719.839	5	-21727.710	5

Supplementary File 1. Summary of the Zika virus sequencing data produced in this study.

Supplementary File 2. Epidemiological data and travelers entering Miami, Florida from January to June, 2016.

Supplementary File 3. Primer sequences used for long-range amplicons and probe sequences for RNA Access targeted enrichment of Zika virus.

Supplementary File 4. Raw MAFFT codon alignment, PhyML tree, BEAST XML file, and BEAST MCC time-structured phylogeny.

References

1. WHO | Zika virus and complications. (2016).
2. Lazear, H. M. & Diamond, M. S. Zika Virus: New Clinical Syndromes and Its Emergence in the Western Hemisphere. *J. Virol.* **90,** 4864–4875 (2016).
3. Likos, A. *et al.* Local Mosquito-Borne Transmission of Zika Virus - Miami-Dade and Broward Counties, Florida, June-August 2016. *MMWR Morb. Mortal. Wkly. Rep.* **65,** 1032–1038 (2016).
4. Hennessey, M., Fischer, M. & Staples, J. E. Zika Virus Spreads to New Areas — Region of the Americas, May 2015–January 2016. *MMWR Morb. Mortal. Wkly. Rep.* **65,** 1–4 (2016).
5. Faria, N. R. *et al.* Zika virus in the Americas: Early epidemiological and genetic findings. *Science* **352,** 345–349 (2016).
6. Regional Zika Epidemiological Update (Americas). *Pan American Health Organization* (2016). Available at: http://www.paho.org/hq/index.php?option=com_content&view=article&id=11599&Itemid=41691&lang=en.
7. Weger-Lucarelli, J. *et al.* Vector Competence of American Mosquitoes for Three Strains of Zika Virus. *PLoS Negl. Trop. Dis.* **10,** e0005101 (2016).
8. Guerbois, M. *et al.* Outbreak of Zika virus infection, Chiapas State, Mexico, 2015, and first confirmed transmission by Aedes aegypti mosquitoes in the Americas. *J. Infect. Dis.* (2016).
9. Ferreira-de-Brito, A. *et al.* First detection of natural infection of Aedes aegypti with Zika virus in Brazil and throughout South America. *Mem. Inst. Oswaldo Cruz* 0 (2016).
10. Chouin-Carneiro, T. *et al.* Differential Susceptibilities of Aedes aegypti and Aedes albopictus from the Americas to Zika Virus. *PLoS Negl. Trop. Dis.* **10,** e0004543 (2016).
11. Kraemer, M. U. G. *et al.* The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *Elife* **4,** e08347 (2015).
12. Ferguson, N. M. *et al.* Countering the Zika epidemic in Latin America. *Science* **353,** 353–354 (2016).
13. Mosquito-Borne Disease Surveillance. *Florida Department of Health* Available at: http://www.floridahealth.gov/diseases-and-conditions/mosquito-borne-diseases/surveillance.html.
14. Teets, F. D. *et al.* Origin of the dengue virus outbreak in Martin County, Florida, USA 2013. *Virol Rep* **1-2,** 2–8 (2014).
15. Graham, A. S. *et al.* Mosquito-associated dengue virus, Key West, Florida, USA, 2010. *Emerg. Infect. Dis.* **17,** 2074–2075 (2011).
16. Kendrick, K. *et al.* Notes from the field: transmission of chikungunya virus in the continental United States—Florida, 2014. *MMWR Morb. Mortal. Wkly. Rep.* **63,** 1137 (2014).
17. Bouri, N. *et al.* Return of epidemic dengue in the United States: implications for the public health practitioner. *Public Health Rep.* **127,** 259–266 (2012).
18. Ramos, M. M. *et al.* Epidemic Dengue and Dengue Hemorrhagic Fever at the Texas–Mexico Border: Results of a Household-based Seroepidemiologic Survey, December 2005. *Am. J. Trop. Med. Hyg.* **78,** 364–369 (2008).
19. Murray, K. O. *et al.* Identification of dengue fever cases in Houston, Texas, with evidence of autochthonous transmission between 2003 and 2005. *Vector Borne Zoonotic Dis.* **13,** 835–845 (2013).
20. Monaghan, A. J. *et al.* On the Seasonal Occurrence and Abundance of the Zika Virus Vector Mosquito Aedes Aegypti in the Contiguous United States. *PLoS Curr.* **8,** (2016).
21. Dzul-Manzanilla, F. *et al.* Evidence of vertical transmission and co-circulation of chikungunya and dengue viruses in field populations of Aedes aegypti (L.) from Guerrero, Mexico. *Trans. R. Soc. Trop. Med. Hyg.* **110,** 141–144 (2016).
22. Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *bioRxiv* 098913 (2017).
23. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29,** 1969–1973 (2012).
24. Robert, M. A. *et al.* Modeling Mosquito-Borne Disease Spread in U.S. Urbanized Areas: The Case of Dengue in Miami. *PLoS One* **11,** e0161365 (2016).
25. Dinh, L., Chowell, G., Mizumoto, K. & Nishiura, H. Estimating the subcritical transmissibility of the Zika outbreak in the State of Florida, USA, 2016. *Theor. Biol. Med. Model.* **13,** 20 (2016).
26. Nelson, B. *et al.* Travel Volume to the United States from Countries and U.S. Territories with Local Zika Virus Transmission. *PLoS Curr.* **8,** (2016).
27. McCarthy, M. First US case of Zika virus infection is identified in Texas. *BMJ* **352,** i212 (2016).
28. Duffy, M. R. *et al.* Zika virus outbreak on Yap Island, Federated States of Micronesia. *N. Engl. J. Med.* **360,** 2536–2543 (2009).
29. Semenza, J. C. *et al.* International dispersal of dengue through air travel: importation risk for Europe. *PLoS Negl. Trop. Dis.* **8,** e3278 (2014).
30. Nunes, M. R. T. *et al.* Air travel is associated with intracontinental spread of dengue virus serotypes 1-3 in Brazil. *PLoS*

Negl. Trop. Dis. **8,** e2769 (2014).

31. Zika virus. *Florida Department of Health* Available at: http://www.floridahealth.gov/diseases-and-conditions/zika-virus/index.html?utm_source=flhealthIndex.

32. Daily Zika update. *Florida Department of Health* Available at: http://www.floridahealth.gov/newsroom/all-articles.html.

33. Zika virus case counts in the US. *The Centers for Disease Control and Prevention* (2016). Available at: http://www.cdc.gov/zika/geo/united-states.html.

34. Rabe, I. B. *et al.* Interim Guidance for Interpretation of Zika Virus Antibody Test Results. *MMWR Morb. Mortal. Wkly. Rep.* **65,** 543–546 (2016).

35. Waggoner, J. J. & Pinsky, B. A. Zika Virus: Diagnostics for an Emerging Pandemic Threat. *J. Clin. Microbiol.* **54,** 860–867 (2016).

36. Lanciotti, R. S. *et al.* Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. *Emerg. Infect. Dis.* **14,** 1232–1239 (2008).

37. Interim Guidance for Zika Virus Testing of Urine - United States, 2016. *MMWR Morb. Mortal. Wkly. Rep.* **65,** 474 (2016).

38. Biggerstaff, B. PooledInfRate, version 4.0. *An Excel Add-In to Compute Infection Rates from Pooled Data. Centers for Disease Control, Fort Collins, CO* (2009).

39. Fauver, J. R. *et al.* Temporal and Spatial Variability of Entomological Risk Indices for West Nile Virus Infection in Northern Colorado: 2006-2013. *J. Med. Entomol.* **53,** 425–434 (2016).

40. National Centers for Environmental Information. *National Oceanic and Atmospheric Administration* Available at: https://www.ncdc.noaa.gov/.

41. Bogoch, I. I. *et al.* Potential for Zika virus introduction and transmission in resource-limited countries in Africa and the Asia-Pacific region: a modelling study. *Lancet Infect. Dis.* (2016). doi:10.1016/S1473-3099(16)30270-5

42. Hwang, W.-H. & He, F. Estimating abundance from presence/absence maps. *Methods Ecol. Evol.* **2,** 550–559 (2011).

43. Corman, V. M. *et al.* Clinical comparison, standardization and optimization of Zika virus molecular detection. *Bull. World Health Organ.* (2016).

44. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

45. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

46. Köster, J. & Rahmann, S. Building and documenting workflows with python-based snakemake. in *OASIcs-OpenAccess Series in Informatics* **26,** (Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2012).

47. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–1649 (2012).

48. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).

49. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27,** 863–864 (2011).

50. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).

51. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).

52. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704 (2003).

53. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22,** 160–174 (1985).

54. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39,** 306–314 (1994).

55. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9,** 772 (2012).

56. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* **2,** vew007 (2016).

57. Shapiro, B., Rambaut, A. & Drummond, A. J. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* **23,** 7–9 (2006).

58. Ferreira, M. A. R. & Suchard, M. A. Bayesian analysis of elapsed times in continuous-time Markov chains. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* **36,** 355–368 (2008).

59. Baele, G. *et al.* Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* **29,** 2157–2167 (2012).

60. Xie, W., Lewis, P. O., Fan, Y., Kuo, L. & Chen, M.-H. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst. Biol.* **60,** 150–160 (2011).

61. Gelman, A. & Meng, X.-L. Simulating Normalizing Constants: From Importance Sampling to Bridge Sampling to Path

Sampling. *Stat. Sci.* **13,** 163–185 (1998).

62. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22,** 1185–1192 (2005).

63. Churcher, T. S. *et al.* Public health. Measuring the path toward malaria elimination. *Science* **344,** 1230–1232 (2014).

64. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438,** 355–359 (2005).

65. Nishiura, H., Yan, P., Sleeman, C. K. & Mode, C. J. Estimating the transmission potential of supercritical processes based on the final size distribution of minor outbreaks. *J. Theor. Biol.* **294,** 48–55 (2012).

66. Perkins, T. A., Scott, T. W., Le Menach, A. & Smith, D. L. Heterogeneity, mixing, and the spatial scales of mosquito-borne pathogen transmission. *PLoS Comput. Biol.* **9,** e1003327 (2013).

67. Kraemer, M. U. G. *et al.* Big city, small world: density, contact rates, and transmission of dengue across Pakistan. *J. R. Soc. Interface* **12,** 20150468 (2015).

68. Zika-Epidemiological Report. *Pan American Health Organization* Available at: http://www.paho.org/hq/index.php?option=com_content&view=article&id=11603&Itemid=41696&lang=en.

69. Zika Cumulative Cases. *Pan American Health Organization* Available at: http://www.paho.org/hq/index.php?option=com_content&view=article&id=12390&Itemid=42090&lang=en.

70. Perkins, T. A., Siraj, A. S., Ruktanonchai, C. W., Kraemer, M. U. G. & Tatem, A. J. Model-based projections of Zika virus infections in childbearing women in the Americas. *Nat Microbiol* **1,** 16126 (2016).

71. Lessler, J. T. *et al.* Times to key events in the course of Zika infection and their implications: a systematic review and pooled analysis. *Bull. World Health Organ.* (2016).