Genomic epidemiology reveals multiple introductions of Zika virus into the United States

Nathan D. Grubaugh1*, Jason T. Ladner2*, Moritz U. G. Kraemer1,4,5*, Gytsis Dudas6*, Amanda L. Tan7*, Karthik GangaRavapu1*, Michael R. Wiley2,8, Stephen White9*, Julien Thézé3*, Diogo M. Magnani10, Karla Prieto2,8, Daniel Reyes2,11, Andrea M. Bingham3, Lauren M. Paul7, Refugio Robles–Sikisaka1, Glenn Oliveira13, Darryl Pronto9, Carolyn M. Barcellona2, Hayden C. Metsky14, Mary Lynn Baniecki8, Kayla G. Barnes4, Bridget Chak4, Christine A. Freije14, Adrianna Gladden–Young14, Andreas Gnirke14, Cynthia Luo14, Bronwyn Maclnnis4, Christian B. Matrang4, Daniel J. Park14, James Qu14, Stephen F. Schaffner14, Christopher Tomkis–Tinch14, Kendra L. West14, Sarah M. Winnicki14, Shirlee Wohlf14, Nathan L. Yozwiak14, Joshua Quick15, Joseph R. Fauver16, Kamran Khan7,18, Shannon E. Brent17, Robert C. Reiner Jr19, Paola N. Lichtenberger20, Michael J. Ricciardi10, Varian K. Bailey10, David I. Watkins10, Marshall R. Cone21, Edgar W. Kopp IV21, Kelly N. Hogan2, Andrew C. Cannons21, Reynald Jean22, Andrew J. Monaghan23, Robert F. Garry24, Nicholas J. Loman15, Nuno R. Faria2, Mario C. Porcelli25, Chalmers Vazquez25, elyse R. nagle2,11, Derek A. T. Cummings26, Danielle Stanek12, Marshall R. Cone21, edgar W. Kopp IV21, Michael R. Wiley2,8

Zika virus (ZIKV) is causing an unprecedented epidemic linked to severe congenital abnormalities1–2. In July 2016, mosquito-borne ZIKV transmission was reported in the continental United States; since then, hundreds of locally acquired infections have been reported in Florida3,4. To gain insights into the timing, source, and likely route(s) of ZIKV introduction, we tracked the virus from its first detection in Florida by sequencing ZIKV genomes from infected patients and Aedes aegypti mosquitoes. We show that at least 4 introductions, but potentially as many as 40, contributed to the outbreak in Florida and that local transmission is likely to have started in the spring of 2016—several months before its initial detection. By analysing surveillance and genetic data, we show that ZIKV moved among transmission zones in Miami. Our analyses show that most introductions were linked to the Caribbean, a finding corroborated by the high incidence rates and traffic volumes from the region into the Miami area. Our study provides an understanding of how ZIKV initiates transmission in new regions.

ZIKV transmission in the Americas was first reported in Brazil in May 20155, although the virus was probably introduced 1–2 years earlier6–8. By January 2016, ZIKV cases had been reported in several South and Central American countries and most islands in the Caribbean9. Like dengue virus (DENV) and chikungunya virus (CHIKV), ZIKV is vectored primarily by Aedes mosquitoes10–13. The establishment of the peridomestic species A. aegypti in the Americas14 has facilitated the establishment of DENV, CHIKV, and now probably ZIKV as endemic in this region15. In the continental United States, transient outbreaks of DENV and CHIKV have been reported in regions of Texas and Florida16–21 with abundant seasonal A. aegypti populations14,22.

The 2016 ZIKV outbreak in Florida generated 256 confirmed ZIKV infections4 (Fig. 1a). While transmission was confirmed across four counties in Florida (Fig. 1b), the outbreak was most intense in Miami-Dade County (241 infections). Although the case location could not always be determined, at least 114 (47%) infections are likely to have been acquired in one of three distinct transmission zones: Wynwood, Miami Beach, and Little River (Fig. 1c, d).

Using mosquito surveillance data, we determined the extent of mosquito-borne ZIKV transmission in Miami. Of the 24,351 mosquitoes collected from June to November 2016, 99.8% were A. aegypti and 8 pools of ≤50 mosquitoes tested positive for ZIKV (Fig. 1c, Extended Data Fig. 1). From these pools, we estimated that approximately 1 out of 1,600 A. aegypti mosquitoes were infected (0.061%, 95% confidence interval (CI) 0.028–0.115, Extended Data Fig. 1a). This is similar to infection rates during DENV and CHIKV outbreaks23. Although we did not detect ZIKV-infected mosquitoes outside Miami Beach (Fig. 1c), we found that the number of human ZIKV cases correlated strongly with A. aegypti abundance within each transmission zone (Spearman $r = 0.61$, Fig. 1d, Extended Data Fig. 1b). This suggests that

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

15 JUNE 2017 | VOL 546 | NATURE | 401

doi:10.1038/nature22400
A. aegypti mosquitoes were the primary mode of transmission and that changes to vector abundance affected human infection rates. We found that the application of insecticides suppressed mosquito populations during periods of intensive use (Extended Data Fig. 1c), and therefore probably contributed to ZIKV clearance.

We sequenced 39 ZIKV genomes from clinical and mosquito samples without cell culture (Supplementary Table 1a). Our ZIKV dataset included 29 genomes from patients with locally acquired infections (Fig. 1d) and 7 from A. aegypti pools (Fig. 1c). We also sequenced three ZIKV genomes from travel-associated cases in Florida. Our dataset included cases from all transmission zones in Miami (Fig. 1d) and represented about 11% of all confirmed locally acquired cases in Florida. We made all sequence data openly available in the NCBI BioProject database (PRJNA342539, PRJNA356429) immediately after data generation.

We reconstructed phylogenetic trees from our ZIKV genomes along with 65 published genomes from other affected regions (Fig. 2, Extended Data Figs 2, 3). We found that the Florida ZIKV genomes formed four distinct lineages (labelled F1–F4 in Fig. 2a), three of which (F1–F3) belonged to the same clade (labelled A in Fig. 2a). We sampled only a single human case each from the F3 and F4 lineages, consistent with limited transmission (Fig. 2a). The other two Florida lineages (F1 and F2) comprised ZIKV genomes from human and mosquito samples within Miami-Dade County (Fig. 2b).

Using time-structured phylogenies, 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 distinct from those in clade A (Fig. 2a). For the two well-supported nodes linking lineages F1 and F2 (labelled B, Fig. 2a) and F1–F3 (A, Fig. 2a), we estimated the time of the most recent common ancestor (tMRCAs) to be during the summer of 2015 (95% highest posterior density (HPD) June–September 2015). Our data displayed a strong clock signal (Extended Data Fig. 2b) and tMRCAs estimates were robust across a range of models (Extended Data Table 1a). Thus, although lineages F1–F3 belong to clade A, any fewer 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 more than a year, we modelled spread under different assumptions of the basic reproductive number ($R_0$). Using the number of locally acquired and travel-associated cases, along with the number of observed genetic lineages, we estimated an $R_0$ between 0.5 and 0.8 in Miami-Dade County (Extended Data 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%, Extended Data Fig. 2c). This is especially true considering the low A. aegypti abundance during the winter months (Extended Data Fig. 1d).

Given the low probability of long-term persistence, we expect that our ZIKV genomes (F1–F4) were the result of at least four introductions. Differences in surveillance practices and a high number of travel-associated cases (Fig. 1a), however, probably mean that unsampled ZIKV introductions also contributed to the outbreak. To estimate the total number of ZIKV introductions, we modelled 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 outbreak (Fig. 2d). The majority of these introductions would be likely to have generated a single secondary case that was undetected in our genetic sampling (Extended Data Fig. 4a). Incorporating under-reporting in a sensitivity analysis increases $R_0$ estimates slightly to 0.7–0.9 (Extended Data Fig. 4f–i).

The two main ZIKV lineages, F1 and F2, included the majority of genomes from Florida (92%, Fig. 2a). Assuming that 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 centred around March–April 2016 (Fig. 2b, 95% HPD January–May 2016). The estimated timing for these introductions corresponds with the presence of suitable A. aegypti populations in Miami-Dade County (Extended Data Fig. 1d) and suggests that ZIKV transmission could have started at least 2 months before its detection in July 2016 (Fig. 1a). The dates of the introductions could be more recent if multiple F1 or F2 lineage viruses arrived independently. However, more than two introductions would be necessary to substantially change our estimates for the timing of the earliest introduction.
To understand transmission dynamics within Miami, we analysed our genomic data together with case data from the Florida Department of Health (Supplementary Table 1a). Although the three ZIKV transmission zones were spatially distinct, they occurred within about 5 km of each other (Fig. 1c) and the ZIKV infections associated with each zone overlapped temporally (Fig. 1d). Our ZIKV genomes with zone assignments all belonged to lineages F1 and F2, but neither of these lineages was confined to a single zone (Fig. 2b). In fact, we detected both F1 and F2 lineage viruses from A. aegypti collected from the same trap 26 days apart (mosquitoes 5 and 8, Fig. 2b). These findings suggest that ZIKV moved among areas of Miami.

Determining the sources and routes of ZIKV introductions could help to mitigate future outbreaks. We found that lineages F1–F3 clustered with ZIKV genomes sequenced from the Dominican Republic and Guadeloupe (Fig. 2, Extended Data Figs 2, 3). By contrast, F4 clustered with genomes from Central America (Fig. 2, Extended Data Figs 2, 3). These findings suggest that whereas ZIKV outbreaks occurred throughout the Americas, the Caribbean islands were the main source of local ZIKV transmission in Florida. Because of severe undersampling of ZIKV genomes, however, we cannot rule out other source areas. 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. We found that the Caribbean islands bore the highest ZIKV incidence rates (Fig. 2b), despite Brazil and Colombia reporting the highest absolute number of cases (January–June 2016; Fig. 3a, Extended Data Fig. 5, Supplementary Table 1b). During the same time period, we estimated that about 3 million travellers arrived from the Caribbean, accounting for 54% of the total traffic into Miami, with the vast majority (about 2.4 million) arriving via cruise ships (Fig. 3b, Extended Data Fig. 6, Supplementary Table 1b). Combining the infection rates with travel capacities, we estimated that about 3–4 million ZIKV-infected travellers arrived from the Caribbean (Fig. 3c, Extended Data Fig. 7a). We also found that the number of travel-associated ZIKV cases correlated strongly with the expected number of importations from the Caribbean (Spearman $r = 0.8$; Fig. 3d, Extended Data Fig. 7b). Finally, we also found that the probability of ZIKV persistence after introduction for different $R_0$ values.

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 coloured according to collection location. The five tips outlined in blue but filled with a different colour indicate ZIKV cases in the US associated with travel (fill colour indicates the probable location of infection). Clade posterior probabilities are indicated by white circles filled with black relative to the level of support. The grey violin plot indicates the 95% HPD interval for the tMRCA for the epidemic in the Americas (AM). 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 tMRCA for the epidemic in the Americas (AM).

c. The probability of ZIKV persistence after introduction for different $R_0$ values. 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 $R_0$ values.
dance. We estimated that Miami and nearby Fort Lauderdale received from regions with ZIKV transmission along with local ZIKV outbreaks in this area, we analysed incoming passenger traffic County (Fig. 1b). To determine whether there is a higher potential for ships. contribution of ZIKV-infected travellers arriving via airlines or cruise their purpose of travel and therefore we cannot determine the specific addition, ZIKV exposure may vary among individuals depending on a major mode of transportation from areas with active transmission. In These findings, however, do not indicate that cruise ships themselves played a key role in the establishment of ZIKV transmission in Florida. ZIKV in the Caribbean, combined with frequent travel, could have is unknown. Together, these findings suggest that a high incidence of recent travel to the Caribbean (Fig. 3e); however, their mode of travel 67% of individuals with travel-associated infections in Florida reported around 72% of traffic (Fig. 4) and that Miami received more air and sea traffic from ZIKV-endemic areas than any other city in the United States (Extended Data Fig. 8). We estimated that, during January–April 2016, A. aegypti abundance was highest in southern Florida (Fig. 4, Extended Data Figs 1d, 8). By June, most of Florida and several cities across the South probably supported high populations of A. aegypti (Extended Data Fig. 8); however, most of this region has not reported local A. aegypti-borne virus transmission for at least 60 years. In fact, the only region outside Florida with local ZIKV transmission is southern Texas, which is also the only other region with recent DENV outbreaks. Therefore, the combination of travellers, mosquito ecology, and human population density is likely to make Miami one of the few places in the continental United States at risk for A. aegypti-borne virus outbreaks.

The extent of ZIKV transmission in Florida was unprecedented, with more reported ZIKV cases in 2016 (256) than DENV cases since 2009 (136)4,16,17. This case difference may be reflected by lower incidence of

Figure 3 | Frequent opportunities for Zika virus introductions into Miami from the Caribbean. a. Reported ZIKV cases per country or territory from January to June 2016, normalized by total population. The number of estimated travellers entering Miami during January–June 2016, by method of travel. c. The number of travellers and reported ZIKV incidence rate for the country or territory of origin were used to estimate the proportion of infected travellers coming from each region with ZIKV in the Americas. d. The observed number of weekly travel-associated ZIKV cases in Florida (black line), plotted with the expected number of ZIKV-infected travellers (as estimated in c) coming from all of the Americas (grey line) and the regional contributions (coloured areas). e. The countries visited by the 1,016 patients with travel-associated ZIKV diagnosed in Florida.

Figure 4 | Southern Florida has a high potential for A. aegypti-borne virus outbreaks. The estimated number of travellers per month (circles) entering Florida cities via flights and cruise ships, plotted with estimated relative A. aegypti abundance. Only cities receiving more than 10,000 passengers per month are shown. Relative A. aegypti abundance for every month is shown in Extended Data Fig. 1d.
endemic DENV than epidemic ZIKV in source countries resulting in fewer DENV importations (reported travel-associated cases since 2009: 654 DENV and 1,016 ZIKV). Given that the majority of ZIKV infections are asymptomatic, the true number of ZIKV cases is likely to have been much higher. Despite this, we estimated that the average R0 was less than 1 and therefore multiple introductions were necessary to give rise to the observed outbreak. The high volume of traffic entering Florida from ZIKV-affected regions, especially the Caribbean, is likely to have provided a substantial supply of ZIKV-infected individuals. Because Florida is unlikely to sustain long-term ZIKV transmission, the potential for future ZIKV outbreaks in this region is dependent upon activity elsewhere. Therefore, we expect that outbreaks in Florida will cycle with ZIKV transmission dynamics in the Americas.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 February; accepted 28 April 2017.

Published online 24 May 2017.

1. Zika virus and complications. World Health Organization http://www.who.int/features/qu/zika/en/ (2016).

2. Laazer, H. M. & Diamond, M. S. Zika virus: new clinical syndromes and its emergence in the Western hemisphere. J. Virol. 8646–8675 (2016).

3. Lima, A. et al. Detection of mosquito-borne transmission of Zika virus—Miami-Dade and Broward Counties, Florida, June–August 2016. MMWR Morb. Mortal. Wkly. Rep. 65, 1032–1036 (2016).

4. Mosquito-Borne Disease Surveillance. Florida Department of Health http://www.floridahealth.gov/diseases-and-conditions/mosquito-borne-diseases-surveillance.html (2017).

5. 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, 55–58 (2016).

6. Faria, N. R. et al. Zika virus in the Americas: early epidemiological and genetic findings. Science 349, 345–349 (2016).

7. Faria, N. R. et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature http://dx.doi.org/10.1038/nature22402 (2017).

8. Metsky, H. C. et al. Zika virus evolution and spread in the Americas. Nature http://dx.doi.org/10.1038/nature22402 (2017).

9. Regional Zika Epidemiological Update (Americas). Pan American Health Organization http://www.paho.org/hq/index.php?option=com_content&view=article&id=115999&Itemid=41691 (2016).

10. Weger-Lucarelli, J. et al. Vector competence of American mosquitoes for three arbovirus vectors—Aedes aegypti—Maido and the BEAUTi and the BEAULi, Gen. Virol. 97, 1–17 (2016).

11. 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).

12. McCarthy, M. First US case of Zika virus infection is identified in Texas. Br. Med. J. 352, i212 (2016).

13. Nelson, B. et al. Travel volume to the United States from countries and U.S. territories with local Zika virus transmission. PLoS Curr. http://dx.doi.org/10.1371/currents.outbreaks.acb0d8c93c5e88825c1a147679531c (2016).

14. Dengue in Puerto Rico. The Centers for Disease Control and Prevention https://www.cdc.gov/dengue/about/inpuerto.html (2017).

15. Pan American Health Organization. Countries and territories with autochthonous transmission of the Zika virus in the Americas reported in 2015–2017. http://www.paho.org/hq/index.php?option=com_content&view=article&id=116038&Itemid=41669 (2016).

16. Duffy, M. R. et al. Zika virus outbreak on Yap Island, Federated States of Micronesia. N. Engl. J. Med. 360, 2536–2543 (2009).

17. Dinh, L., Chowell, G., Mizumoto, K. & Nishiura, H. Estimating the subclinical transmissibility of the Zika virus outbreak in the State of Florida, USA, 2016. Theor. Biol. Med. Model. 13, 20 (2016).

18. 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, e2876 (2014).

supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Weger-Lucarelli, G. Ebel, C. Moore, B. Alto, G. Donatti, and S. Taylor for discussions; E. Spencer for IRB and logistics support; M. Pichler for sequencing assistance; and G. Schroth and S. Gross for designing and providing enrichment probes. N.D.G. is supported by NIH training grant 5T32AI007244-33. D.G. is supported by the Mahan Postdoctoral Fellowship from the Computational Biology Program at Fred Hutch. K.G.B. is supported by the ASTMH Shape Fellowship. N.R.F. is funded by a Sir Henry Dale Fellowship (Wellcome Trust/Royal Society grant 204311/2/16/2). D.A.T.C. was supported by US NIH MIDAS program (U54-GM088491) and CDC Cooperative Agreement U01CK000510. A.R. is supported by EU Seventh Framework Programme (FP7/2007-2013) grant 278433-PREDIVMICS, ERC grant 260864, Horizon 2020 grant 643476-COMARE, P.C.S. is supported by Marc and Lynne Benioff; NIH NIAID U19AI10818; Howard Hughes Medical Institute; AWS Cloud Credits for Research; Broad Institute BroadNext10 program (also A.G.). T.B. is a Pew Biomedical Scholar and is supported by NIH R35 GM119774-01. O.G.P. received funding from EU Seventh Framework Programme (FP7/2007-2013)/ERC number 614725-PATHPHYLODYN and the USAID Emerging Pandemic Threats Program-2 PREDICT-2 (Cooperative Agreement No. OA-AAA-A-14-00102). S.I. and S.F.M are supported by NIH NIAID U19AI10818. M.R.C. was supported by the US Armed Forces Research Institute of Environmental medicine, North Franklin, Connecticut. K.G.A. is a Pew Biomedical Scholar, and is supported by NIH NCATS CTSA UL1TR001114. K.G.B., B.C., C.A.F., A.G.-Y., A.G., C.L., B.M., C.B.M., D.J.P., J.Q.U, S.F.S., C.T.-T., N.D.G., M.R.W., K.P., D.R., R.R.-S., G.O., and E.R.N. provided data, reagents, or materials; A.M.B., L.M.P., D.P., C.M.B., P.N.L., M.J.R., V.K.B., D.I.W., M.R.C., E.W.K., K.N.H., G.P., and K.G.A. collected samples: A.L.T., S.W., D.M.M., T.B., O.G.P., S.I., G.P., and K.G.A. contributed to the study design, and the other anonymous reviewer(s) for their contribution to the peer review process.

Author Contributions All authors contributed to the study design. Designed the experiments: N.D.G., J.T.L., G.D., M.U.G.K., D.A.T.C., P.C.S., L.D.G., S.F.M., T.B., O.G.P., S.I., and K.G.A. Collected samples: A.M.B., L.M.P., D.P., C.M.B., P.N.L., M.J.R., K.V.B., D.I.W., M.R.C., E.W.K., K.N.H., A.C.C., R.J., M.C.P., C.V., D.S., L.D.G., S.F.M., and S.I. Obtained the sequencing data: N.D.G., M.R.W., K.P., D.R., R.R.-S., G.D., and G.E.R. Provided reagents, antibodies, or samples: N.D.G., J.T.L., G.D., M.U.G.K., K.G., M.R.W., R.R.-S., G.O., H.C.M., M.L.B., K.G.B., C.B., C.A.F., A.G.-Y., A.G., C.L., B.M., C.B.M., D.J.P., J.Q., S.F.S., C.T.T., K.L.W., S.M.W., S.W., N.L.Y., J.Q.J., J.R.F., K.K., S.E.B., A.M.G., R.F.J., N.L.L., M.C.P., C.V., P.C.S., S.F.M., and S.I. Analysed the data: N.D.G., J.T.L., G.D., M.U.G.K., K.G., T.B., O.G.P., S.I., and G.P. Wrote the manuscript: D.G., M.U.G.K., J.T.L., S.F.S., A.R., T.B., O.G.P., S.I., and G.P. Manuscript reviewed: D.G., M.U.G.K., J.T.L., S.F.S., A.R., T.B., O.G.P., S.I., and G.P. Manuscript reviewed.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to K.G.A. (andersen@scripps.edu) or O.G.P. (gustavo.f.palacios.ctr@mail.mil).

Reviewer Information Nature thanks K. St George, A. Wider-Smith, M. Worobey and the other anonymous reviewer(s) for their contribution to the peer review of this work.
METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Ethical statement. This work was evaluated and approved by the relevant Institutional Review Boards (IRBs) or Ethics Review Committees at The Scripps Research Institute (TSRI) and the US Army Medical Research Institute of Infectious Diseases (USAMRIID) Office of Human Use and Ethics. This work was conducted as part of the public health response in Florida and samples were collected under a waiver of consent granted by the Florida Department of Health (DOH) Human Research Protection Program. The work received a non-human subjects research designation (category 4 exemption) by the Florida DOH because this research was performed with leftover clinical diagnostic samples involving no more than minimal risk. All samples were de-identified before receipt by the study investigators.

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 system5. 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 website38 and daily updates35. International travel-associated ZIKV case counts in the United States (outside Florida) were obtained from the CDC36. 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 five or more times by ZIKV-infected travellers diagnosed in Florida are shown. Countries with fewer than five visits were aggregated into an “other” category by region (that is, Caribbean, South America, or Central America).

Clinical sample collection and RNA extraction. Clinical samples from locally acquired ZIKV infections were collected between 22 June and 11 October 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 guidelines37–39. Clinical specimens (whole blood, serum, saliva, or urine) submitted for analysis were refrigerated or frozen at or below −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 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. Tenfold dilutions of the ZIKV RNA transcripts (5 μl per reaction) were used to create a standard curve for quantification of ZIKV GE per μl RNA. The lower limits of quantification are 4 GE per μl RNA, or at a cycle threshold of ~36.

ZIKV GE were quantified at USAMRIID using the University of Bonn ZIKV envelope protein (Bonn E) qRT–PCR assay44. RNA standards were transcribed using an ampiclon 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 it included nt 1–4,348, which covers the 5′ UTR, C, prM, 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 described44. Tenfold dilutions of the ZIKV RNA transcripts (5 μl per reaction) were used to create a standard curve for quantification of ZIKV GE per μl RNA.

Amplicon-based Zika virus sequencing. ZIKV sequencing at TSRI was performed using an ampiclon-based approach using the ZikaAsian V1 scheme, as described44. This approach is similar to ‘RNA jackhammering’ to sequence low-quality viral samples described in ref. 45. In brief, cDNA was reverse-transcribed from 5 μl RNA using SuperScript IV (Invitrogen). ZIKV cDNA (2.5 μl per 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 X2 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 (ref. 46). 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 statistics47. Snakemake was used as the workflow management system48. 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 (ref. 49) before generating consensus sequences. A minimum of 3 × read-depth coverage, in support of the consensus, was required to make a base call.

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 indexing, with no overlapping indices, was used. The probes were designed to cover the entire ZIKV genome and to encompass the genetic diversity present on GenBank on 14 January 2016. In total, 26 ZIKV sequences were used during probe design (Supplementary Table 2b). 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 V3 kit (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 v1.9.6 (ref. 50), and low-quality reads or bases were filtered using Prinseq-lite v0.20.3 (ref. 51). Reads were aligned to a reference genome (GenBank: KX197192.1) using Bowtie2 v2.0.6 (ref. 52). Duplicates were removed with Picard (http://broadinstitute.github.io/picard), and a new consensus abundance in major US cities presented in Extended Data Fig. 8 was estimated as previously described22.

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 (nucleotides (nt) 8,651–9,498) using the T7 forward primer (5′-TAATACGACTCACTATAGGGAGGAGTCCGAGGCTCTGTTAAAACCC-3′), reverse primer (5′-AGTGCAACATGGTCGCTTC-3′), and the T7 Megascript kit (Ambion). For qRT–PCR, primers and probes targeting the NS5 region (nt 8,650–9,498) were designed using the ZIKV isolate PRVABC29 (GenBank: KU501215): forward primer (5′-AGTGCCAGAGGTTGTTGTA-3′), reverse primer (5′-CTTAGCTCTAGCCCATGT-3′), and FAM-fluorescent probe (5′-GGCGAGCCGGCCATCTGGT-3′). 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. Tenfold dilutions of the ZIKV RNA transcripts (2 μl per reaction) were used to create a standard curve for quantification of ZIKV GE per μl RNA. The lower limits of quantification are 4 GE per μl RNA, or at a cycle threshold of ~36.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
was generated using a combination of Samtools v0.1.18 (ref. 47) and custom scripts (https://github.com/itladner/Scripts/blob/master/reference-based_assembly_consortium.fasta.py). Only bases with Phred quality score ≥20 were used 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 (that is, called as ‘N’).

Validation and comparison of sequencing methods. 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 custom-targeted towards ZIKV on the Ion S5 platform; and (3) sequencing 5 × 2,150–2,400-bp ZIKV amplicons on the MiSeq. For 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 Table 2c) were kindly provided by S. O’Connor, D. Dudly, D. O’Connor, and D. Gellerup (AIDS Vaccine Research Laboratory, University of Washington, Madison). Each fragment was amplified individually by PCR using the cDNA generated above, Q5 DNA High-fidelity Polymerase, and the following thermocycle conditions: 55°C for 30 min, 94°C for 2 min, 35 cycles of 94°C for 15 s, 56°C for 30 s, and 68°C for 3.5 min, 68°C for 10 min, and held at 4°C until use. Each PCR product was purified using Agencourt AMPure XP beads, sheared to 300–400-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 (Extended Data Table 2). There were, however, some mismatches in the 5′ and 3′ UTRs (where the genomic RNA is heavily structured), probably as a result of PCR bias and decreased coverage depth.

At least 95% of the ZIKV genome was covered from samples with as low as 4 and 9 GE per μl RNA from the amplicon and enrichment approaches, respectively. These results are similar to our previously determined clinical range of 10–16 ZIKV GE per μl RNA to achieve at least 95% genome coverage using our amplicon-based approach44. On average, the amplicon-based sequencing approach covered 97% of the ZIKV genome (≥3 × read-depth) and the targeted enrichment approach covered 82% of the ZIKV genome from clinical samples (Supplementary Table 2d).

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 in December 2016. Public sequences (n = 65) were codon-aligned together with ZIKV genomes generated in this study (n = 39) using Mafft v753 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).

To determine the temporal signal of the sequence dataset, a maximum likelihood (ML) phylogeny was first reconstructed with PhyML 54 using the general time-reversible (GTR) nucleotide substitution model and gamma distributed rates amongst sites52 (Supplementary File 1), which was identified as the best fitting model of sequence evolution by ModelTest53. The ML tree was rooted with a few outgroup sequences (BACR001, ZIKV sequences from Miami-Dade County). We approached this by sampling 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 (Extended Data Fig. 4b). We also calculated the likelihood of observing 241 local cases in the total outbreak (Extended Data Fig. 4c), finding that the MLE of R0 lies between 0.35 and 0.55. As a sensitivity analysis, we additionally modelled introductions with the assumption that only 50% of travellers were infectious at time of arrival into Miami-Dade County, resulting in an MLE of R0 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 three 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 (j1, j2, ..., jN). We then sampled 27 cases without replacement from (j1, j2, ..., jN) following a hypergeometric distribution and recorded the number of distinct clusters drawn in the sample. We found that higher values of R0 resulted in fewer distinct clusters within the sample of 27 genomes (Extended Data Fig. 4d). We additionally calculated the likelihood of sampling three distinct genetic clusters in 27 genomes (Extended Data Fig. 4e), finding an MLE estimate of R0 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 Eq. (1) in each case. In this way, we found an MLE estimate of R0 of 0.5–0.9.

Using the overlap of estimates of R0 from local case counts (0.35–0.8) and genetic clusters (0.5–0.9), we arrived at a 95% uncertainty range of R0 of 0.5–0.8. As an additional sensitivity analysis, we incorporated under-reporting in which either 50% of travel-associated cases and 25% of local cases are reported or in which 10% of travel-associated cases and 5% of local cases are reported. We found that differential reporting of travel and local cases resulted in increased mean R0 estimates when comparing counts of travel-associated to local cases (Extended Data Fig. 4f, g). Additionally, under-reporting increased estimates of R0 from the sampling analysis (Extended Data Fig. 4h, i). Thus, moderate under-reporting is consistent with R0 estimates of ~0.8.

We additionally performed birth—death stochastic simulations assuming a serial interval with mean of 20 days35. We recorded the number of stochastic simulations still persisting after a particular number of days for different values of R0 (Fig. 2c).

Zika virus incidence rates. Weekly suspected and confirmed ZIKV case counts from countries and territories within the Americas with local transmission (1 January to 18 September 2016) were obtained from the Pan American Health Organization (PAHO)56. In most cases, the weekly case numbers per country were reported only in bar graphs. We contacted PAHO multiple times with the hope of gaining access to the raw data included in the bar graphs, but our requests were unfortunately denied. Therefore we used WebPlotDigitizer v3.10 (http://arohatgi.info/WebPlotDigitizer) to estimate the numbers. We compared the actual ZIKV case number reported in Ecuador (the only country with available raw data and reported cases over 10 per week) to our estimates from the PAHO bar graphs and found that the WebPlotDigitizer was ~99% accurate (Extended Data Fig. 5a, b).
Country and territory total population sizes to calculate weekly and monthly ZIKV incidence rates were also obtained from PAHO\textsuperscript{64}. Incidence rates calculated from countries and territories in the Americas during January–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.

**Airline and cruise ship traffic.** To investigate whether the transmission of ZIKV in Florida coincided with travel patterns from ZIKV endemic regions, we obtained the number of passengers arriving at airports in Florida via commercial air travel. We collated data for flights arriving at all commercial airports in Florida 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 Table 1b). 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.\textsuperscript{39} previously reported flight data from 33 countries with ZIKV transmission entering major US airports from October 2014 to 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.cruiseport.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.cruisecruiser.com. Every country or territory with ZIKV transmission visited by a cruise ship 10 days (the approximate mean time to ZIKV clearance in human blood (that is, the infectious period))\textsuperscript{65} before arrival was counted as contributing the ship’s capacity worth of passengers to Miami to the month of arrival (Supplementary Table 1b). While the air traffic was based on the reported number of travellers, we estimated the sea traffic by ship capacity. Lee and Ramdeen\textsuperscript{37} reported that the average occupancy of cruise ships travelling to the Caribbean Islands exceeded 100% in 2011, and according to the Florida-Caribbean Cruise Association\textsuperscript{77}, it remained >100% in 2015. Occupancy data for 2016 was not available at the time of publication, but we assumed that it was also near 100%.

**Expected number of travellers infected with Zika virus.** We estimated the expected number of travellers entering Miami who were infected with ZIKV ($\lambda$) 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 travellers from each country or territory with ZIKV transmission by region and travel method (flights or cruises). The number of ZIKV cases reported by each country are likely to be underestimates, in part because the majority of ZIKV infections are asymptomatic\textsuperscript{2,31}. We normalized some of the potential reporting variances between countries by reporting the data as the relative proportion of infected travellers (Fig. 3c, Extended Data Fig. 7a) and as the absolute number of infected travellers (Fig. 3d, Extended Data Fig. 7b, Supplementary Table 1b) from each region. We also accounted for potential reporting biases with incidence rates by using ZIKV attack rates (that is, proportion infected before epidemic burnout) to estimate peak transmission intensity. Attack rates were calculated using a susceptible–infected–recovered (SIR) transmission model with seroprevalence studies and environmental factors as described\textsuperscript{33}. Using attack rates as an estimate of infection likelihood, we predict that ~60% of the infected travellers entering Miami came from the Caribbean (Extended Data 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 Table 1b.

**Maps.** The maps presented in our figures were generated using Matplotlib\textsuperscript{59} and ESRI basemaps (www.esri.com/data/basemaps). The software and basemaps are open source and freely available to anyone.

**Data availability.** All ZIKV sequencing data are available under NCBI BioProjects PRJNA35359 and PRJNA356429. Individual sample GenBank access numbers are listed in Supplementary Table 1a. All other data are available in the Extended Data or Supplementary Information, or upon request.

34. Zika virus. Florida Department of Health http://www.floridahealth.gov/diseases-and-conditions/zika-virus/index.html?utm_source=flhealthlink (20017).

35. Daily Zika update. Florida Department of Health http://www.floridahealth.gov/newsroom/all-articles.html (2017).

36. Zika virus case counts in the US. The Centers for Disease Control and Prevention http://www.cdc.gov/zika/geo/united-states.html (2016).
71. 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).
72. Rocklöv, J. et al. Assessing seasonal risks for the introduction and mosquito-borne spread of Zika virus in Europe. EBioMedicine 9, 250–256 (2016).
73. National Direction of Epidemiological Surveillance for Vector Transmitted Diseases (Spanish) Ministry of Public Health, Ecuador. http://www.salud.gob.ec/wp-content/uploads/2015/12/GACETA-ZIKA_SE5corregido.pdf (2017).
74. Zika Cumulative Cases. Pan American Health Organization http://www.paho.org/hq/index.php?option=com_content&view=article&id=12390&Itemid=42090&lang=en (2016).
75. 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. 94, 841–849 (2016).
76. Lee, S. & Ramdeen, C. Cruise ship itineraries and occupancy rates. Tourism Manage. 34, 236–237 (2013).
77. FCCA Research & Statistics. Florida-Caribbean Cruise Association http://www.f-cca.com/research.html (2017).
78. Alex Perkins, T., 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).
79. Hunter, J. D. Matplotlib: A 2D graphics environment. Comput. Sci. Eng. 9, 90–95 (2007).
Extended Data Figure 1 | Miami-Dade mosquito surveillance and relative *A. aegypti* abundance. a, Mosquito surveillance data reported from 21 June to 28 November 2016 was used to evaluate the risk of ZIKV infection from mosquito-borne transmission in Miami. A total of 24,306 *A. aegypti* and 45 *A. 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 the 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 an MLE. None of the *A. albopictus* pools contained ZIKV RNA. b, The number of weekly ZIKV cases (based on symptom onset) was correlated with mean *A. 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 Miami-Dade Mosquito Control in Wynwood (left) and Miami Beach (right) was overlaid with *A. aegypti* abundance per trap night to demonstrate that intense usage of insecticides may have helped to reduce local mosquito populations. d, Relative *A. 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.
Extended Data Figure 2 | Maximum likelihood tree and root-to-tip regression of Zika virus genomes from Pacific islands and the epidemic in the Americas. **a.** Maximum likelihood tree of publicly available ZIKV sequences and sequences generated in this study (n = 104). Tips are coloured by location and labels in bold indicate sequences generated in this study. Florida clusters F1–F4 are indicated by vertical lines to the right of the tree. Bootstrap support values are shown at key nodes. All other support values can be found in Supplementary File 1. **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 $1.10 \times 10^{-3}$ nucleotide substitutions per site per year (subs/site/yr). This is consistent with BEAST analyses using a relaxed molecular clock and a Bayesian Skyline tree prior, the best performing combination of clock and demographic model according to MLEs (Extended Data Table 1c), which estimated an evolutionary rate of $1.21 \times 10^{-3}$ (95% highest posterior density: $1.01–1.43 \times 10^{-3}$) substitutions per site per year (Extended Data Table 1a). These values are in agreement with previous estimates based on ZIKV genomes from Brazil.© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Extended Data 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 the Americas (n = 104). Clades and nodes are coloured according to origin location. Clade posterior probabilities are indicated by white circles filled with black relative to the support. A posterior probability of 1 fills the entire circle black. The grey violin plot indicates the 95% 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 Data Table 1, 95% HPD: August 2013–January 2014), which is consistent with previous analysis based on ZIKV genomes from Brazil6.
Extended Data Figure 4 | See next page for caption.
Extended Data Figure 4 | Estimation of basic reproductive number and number of introductions in Miami-Dade County. 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 (that is, travel-associated cases diagnosed in Miami-Dade County) for different values of $R_0$ and for different assumptions of proportion of infectious travellers. c, log likelihood of observing 241 local cases in Miami-Dade County with 320 introduction events for different values of $R_0$ along with 95% MLE bounds on $R_0$. d, Mean and 95% uncertainty interval for total number of distinct phylogenetic clusters observed in 27 sequenced ZIKV genomes from human cases diagnosed in Miami-Dade County for different values of $R_0$ and for different assumptions of sampling bias, from $\alpha = 1$ (no sampling bias) to $\alpha = 2$ (skewed towards preferentially sampling larger clusters). e, log likelihood of observing three clusters (that is, ZIKV lineages F1, F2, and F4, Fig. 2a) in 27 sequenced cases for different values of $R_0$ along with 95% MLE bounds on $R_0$. f, Mean and 95% CI for total number of local cases caused by 320 observed travel-associated cases with travel-associated versus local reporting rates of 50%/25% and 10%/5%. This assumes that 50% of travellers are infectious. g, log likelihood of observing 241 local cases with 320 introduction events for different values of $R_0$ along with 95% MLE bounds on $R_0$ with travel-associated versus local reporting rates of 50%/25% and 10%/5%. h, 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 assumptions of local reporting rate of 5% and 25%. This assumes preferential sampling ($\alpha = 2$). i, log likelihood of observing three clusters in 27 sequenced cases for different values of $R_0$ along with 95% MLE bounds on $R_0$ with local reporting rates of 5% and 25%. At 5% local reporting rate, none of the 100,000 replicates for all $R_0$ values showed three clusters.
Extended Data Figure 5 | Weekly reported Zika virus case numbers and incidence rates in the Americas. a, Most ZIKV case numbers reported by PAHO were available only as bar graphs (raw data were not made available to us at the time of request). Therefore we used WebPlotDigitizer to estimate the weekly case numbers from the PAHO bar graphs. ZIKV cases reported from Ecuador was the only dataset to include a link to the actual case numbers that also had more than 10 cases per week. To validate the WebPlotDigitizer-derived values, we compared the weekly reported case numbers from Ecuador to our estimates. b, The reported and estimated case numbers were strongly correlated (Spearman \( r = 0.9981 \)). WebPlotDigitizer was used to estimate the ZIKV case numbers for all subsequent analysis. c, d, ZIKV cases (suspected and confirmed; c) and incidence rates (normalized per 100,000 population; d) are shown for each country or territory with available data per epidemiological week from 1 January to 18 September 2016. e, Each country or territory with available data is coloured by its reported ZIKV incidence rate from January to June 2016 (the time frame for analysis of ZIKV introductions into Florida).
Extended Data Figure 6 | Cruise and flight traffic entering Miami from regions with Zika virus transmission. a, b. The estimated number of passengers entering Miami, by cruises (a) or flights (b), from each country or territory in the Americas with ZIKV transmission per month (left). The centre map and inset show the cumulative numbers of travellers 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 (that is, cruises and flights) entering Miami per month.
Extended Data Figure 7 | Expected number of ZIKV-infected travellers from the Caribbean correlated with the total observed number of travel-associated infections. a, To account for potential biases in ZIKV reporting accuracy, we also estimated the proportion of infected travellers using projected ZIKV attack rates (that is, predicted proportion of population infected before epidemic burnout). About 60% of the infected travellers 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 was estimated by the number of travellers coming into Miami from each country or territory (travel capacity) and the in-country or in-territory infection likelihood (incidence rate per person) per week. The expected travel cases were summed from all of the Americas (left), Caribbean (left centre), South America (right centre), 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 indicate a negative correlation between the number of expected and observed travel cases.
Extended Data Figure 8 | Greater early season potential for Zika virus introductions into Miami. The monthly cruise ship and airline capacity from countries or territories with ZIKV transmission for the major US travel hubs (shown as circle diameter) with monthly potential *A. aegypti* abundance (circle colour), as previously estimated. The abundance ranges were chosen with respect to the May–October Miami mean: None to low (<2%); Low to moderate (2–25%); Moderate to high (25–75%); and High (>75%). Mosquito-borne transmission is unlikely in the 'None to low' range. Cruise capacities from Houston and Galveston, Texas were combined.
## Extended Data Table 1 | Evolutionary analyses and model selection

### a

| Model combination     | Clade AM tMRCA Mean | Clade AM tMRCA Lower 95% HPD | Clade AM tMRCA Higher 95% HPD | Clade A tMRCA Mean | Clade A tMRCA Lower 95% HPD | Clade A tMRCA Higher 95% HPD | Clade B tMRCA Mean | Clade B tMRCA Lower 95% HPD | Clade B tMRCA Higher 95% HPD |
|-----------------------|---------------------|-----------------------------|-------------------------------|-------------------|-----------------------------|-------------------------------|-------------------|-----------------------------|-------------------------------|
| Strict, Constant      | 2013.87             | 2013.65                     | 2014.09                       | 2015.59           | 2015.43                     | 2015.76                       | 2015.70           | 2015.52                     | 2015.86                       |
| Strict, Exponential   | 2013.90             | 2013.69                     | 2014.09                       | 2015.55           | 2015.36                     | 2015.72                       | 2015.66           | 2015.47                     | 2015.83                       |
| Strict, Bayesian Skyline | 2013.92          | 2013.70                     | 2014.12                       | 2015.56           | 2015.38                     | 2015.74                       | 2015.69           | 2015.50                     | 2015.86                       |
| UCLN, Constant        | 2013.95             | 2013.66                     | 2014.21                       | 2015.70           | 2015.49                     | 2015.88                       | 2015.82           | 2015.63                     | 2016.03                       |
| UCLN, Exponential     | 2013.95             | 2013.70                     | 2014.19                       | 2015.62           | 2015.39                     | 2015.81                       | 2015.73           | 2015.51                     | 2015.93                       |
| UCLN, Bayesian Skyline | 2013.97          | 2013.72                     | 2014.23                       | 2015.62           | 2015.40                     | 2015.81                       | 2015.76           | 2015.54                     | 2015.94                       |

### b

| Model combination     | Path Sampling | Ranking | Stepping Stone | Ranking |
|-----------------------|--------------|---------|----------------|---------|
| Strict, Constant      | -20379.871   | 5       | -20384.201     | 5       |
| Strict, Exponential   | -20375.747   | 4       | -20379.272     | 4       |
| Strict, Bayesian Skyline | -20358.676  | 2       | -20363.748     | 2       |
| UCLN, Constant        | -20381.228   | 6       | -20385.807     | 6       |
| UCLN, Exponential     | -20370.555   | 3       | -20374.418     | 3       |
| UCLN, Bayesian Skyline | -20354.940  | 1       | -20361.218     | 1       |

**a.** Time of the most recent common ancestor and evolutionary rate. **b.** Model selection to infer time-structured phylogenies. HPD, highest posterior density. Dates listed as proportion of days elapsed with a year. Clades refer to Fig. 2a.
Extended Data Table 2  |  Validation of sequencing results

| Sample   | Amplicon method | NGS platform | 3' UTR | CDS    | 5' UTR |
|----------|-----------------|--------------|--------|--------|--------|
| FL01M    | 35 × 400 bp     | Ion S5       | 1/80   | 0/10272| 7/252  |
|          | 75 × ~200 bp    | Ion S5       | 2/75   | 0/10272| 4/205  |
|          | 5 × ~2200 bp    | MiSeq        | 0/80   | 0/10272| 0/32   |
| FL03M    | 35 × 400 bp     | Ion S5       | 3/87   | 0/10272| 20/252 |
|          | 75 × ~200 bp    | Ion S5       | 4/78   | 0/10272| 5/198  |
|          | 5 × ~2200 bp    | MiSeq        | 0/82   | 0/10272| 0/32   |

NGS, next-generation sequencing; UTR, untranslated region; CDS, coding sequence.

aCompared to the consensus genomes generated by sequencing 35 × 400-bp amplicons on the MiSeq.

bAmplicons produced using Ion AmpliSeq and 875 custom ZIKV primers.