LETTER

doi:10.1038/nature22401

Establishment and cryptic transmission of Zika virus in Brazil and the Americas

N. R. Faria[1,2]*, J. Quick[3]*, I. M. Claro[4]*, J. Thézé[1]*, J. G. de Jesus[5]*, M. Giovanetti[5,6]*, M. U. G. Kraemer[1,7,8]*, S. C. Hill[1]*, A. Black[9,10]*, A. C. da Costa[3], L. C. Franco[2], S. P. Silva[2], C.-H. Wu[11], J. Raghwani[1], S. Cauchemez[12,13], L. du Plessis[1], M. P. Verotti[14], W. K. de Oliveira[15,16], E. H. Carmo[17], G. E. Coelho[18,19], A. C. F. S. Santelli[18,20], L. C. Vinhal[18], C. M. Henriques[17], J. T. Simpson[21], M. Loose[22], K. G. Andersen[23], N. D. Grubaugh[23], S. Somasekar[24], C. Y. Chiu[24], J. E. Muñoz-Medina[25], C. R. Gonzalez-Bonilla[25], C. F. Arias[26], L. L. Lewis-Ximenez[27], S. A. Baylis[28], A. O. Chieppe[29], S. F. Aguiar[29], C. A. Fernandes[29], P. S. Lemos[2], B. L. S. Nascimento[2], H. A. O. Monteiro[2], I. C. Siqueira[5], M. G. de Queiroz[30], T. R. de Souza[30,31], J. F. Bezerra[30,32], M. R. Lemos[33], G. F. Pereira[33], D. Loudal[33], L. C. Moura[33], R. Dhalia[34], R. F. França[34], T. Magalhães[34,35], E. T. Marques Jr[34], T. Jaenisch[37], G. L. Wallau[34], M. C. de Lima[38], V. Nascimento[38], E. M. de Cerqueira[38], M. M. de Lima[39], D. L. Mascarenhas[40], J. P. Moura Neto[41], A. S. Levin[4], T. R. Tozetto-Mendoza[4], S. N. Fonseca[42], M. C. Mendes-Correa[4], F. P. Milagres[43], A. Segurado[4], E. C. Holmes[44], A. Rambaut[45,46], T. Bedford[7], M. R. T. Nunes[2,47]§, E. C. Sabino[4]§, L. C. J. Alcantara[5]§, N. J. Loman[3]§ & O. G. Pybus[1,48]§

Transmission of Zika virus (ZIKV) in the Americas was first confirmed in May 2015 in northeast Brazil[1]. Brazil has had the highest number of reported ZIKV cases worldwide (more than 200,000 by 24 December 2016[2]) and the most cases associated with microcephaly and other birth defects (2,366 confirmed by 31 December 2016[2]). Since the initial detection of ZIKV in Brazil, more than 45 countries in the Americas have reported local ZIKV transmission, with 24 of these reporting severe ZIKV-associated disease[3]. However, the origin and epidemic history of ZIKV in Brazil and the Americas remain poorly understood, despite the value of this information for interpreting observed trends in reported microcephaly. Here we address this issue by generating 54 complete or partial ZIKV genomes, mostly from Brazil, and reporting data generated by a mobile genomics laboratory that travelled across northeast Brazil in 2016. One sequence represents the earliest confirmed ZIKV infection in Brazil. Analyses of viral genomes with ecological and epidemiological data yield an estimate that ZIKV was present in northeast Brazil by February 2014 and is likely to have disseminated from there, nationally and internationally, before the first detection of ZIKV in the Americas. Estimated dates for the international spread of ZIKV from Brazil indicate the duration of pre-detection cryptic transmission in recipient regions. The role of northeast Brazil in the establishment of ZIKV in the Americas is further supported by geographic analysis of ZIKV transmission potential and by estimates of the basic reproduction number of the virus.

Previous phylogenetic analyses have indicated that the ZIKV epidemic was caused by the introduction of an Asian genotype lineage into the Americas around late 2013, at least one year before its detection there[4]. An estimated 100 million people in the Americas are predicted to be at risk of acquiring ZIKV once the epidemic has reached its full extent[5]. However, little is known about the genetic diversity and transmission history of the virus in Brazil[6]. Reconstructing the spread of ZIKV from case reports alone is challenging because symptoms (typically fever, headache, joint pain, rashes, and conjunctivitis) overlap with those caused by co-circulating arthropod-borne viruses[7] and owing to a lack of nationwide ZIKV-specific surveillance in Brazil before 2016.

We undertook a collaborative investigation of the molecular epidemiology of ZIKV in Brazil, including results from a mobile genomics laboratory that travelled through northeast Brazil during June 2016 (the ZiBRA project; http://www.zibraproject.org). Of five regions of Brazil (Fig. 1a), the northeast region has the most notified ZIKV cases (40% of Brazilian cases) and the most confirmed microcephaly cases (76% of Brazilian cases, as of 31 December 2016[2]), raising questions about why the region has been so severely affected[8]. Furthermore, northeast Brazil is the most populous region of Brazil that also has potential for year-round ZIKV transmission[9]. With support from the Brazilian

[1]Department of Zoology, University of Oxford, Oxford OX1 3SY, UK. [2]Evandro Chagas Institute, Ministry of Health, Ananindeua, Brazil. [3]Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK. [4]Department of Infectious Disease, School of Medicine & Institute of Tropical Medicine, University of São Paulo, São Paulo, Brazil. [5]Fundação Oswaldo Cruz (FIOCRUZ), Salvador, Bahia, Brazil. [6]University of Rome Tor Vergata, Rome, Italy. [7]Harvard Medical School, Boston, Massachusetts, USA. [8]Boston Children's Hospital, Boston, Massachusetts, USA. [9]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [10]Department of Epidemiology, University of Washington, Seattle, Washington, USA. [11]Department of Statistics, University of Oxford, Oxford OX1 3LB, UK. [12]Mathematical Modelling of Infectious Diseases and Center of Bioinformatics, Biostatistics and Integrative Biology, Institut Pasteur, Paris, France. [13]Centre National de la Recherche Scientifique, URA3012, Paris, France. [14]Coordenação dos Laboratórios de Saúde (CGLAB/DEVIT/SVS), Ministry of Health, Brasília, Brazil. [15]Coordenação Geral de Vigilância e Resposta às Emergências em Saúde Pública (CGVR/DEVIT), Ministry of Health, Brasília, Brazil. [16]Center of Data and Knowledge Integration for Health (CIDACS), Fundação Oswaldo Cruz (FIOCRUZ), Salvador, Brazil. [17]Departamento de Vigilância das Doenças Transmissíveis, Ministry of Health, Brasilia, Brazil. [18]Coordenação Geral dos Programas de Controle e Prevenção da Malária e das Doenças Transmitidas pelo *Aedes*, Ministry of Health, Brasília, Brazil. [19]Pan American Health Organization (PAHO), Buenos Aires, Argentina. [20]Fundação Oswaldo Cruz (FIOCRUZ), Rio de Janeiro, Brazil. [21]Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [22]University of Nottingham, Nottingham, UK. [23]Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, California 92037, USA. [24]Departments of Laboratory Medicine and Medicine & Infectious Diseases, University of California, San Francisco, California, USA. [25]División de Laboratorios de Vigilancia e Investigación Epidemiológica, Instituto Mexicano del Seguro Social, Ciudad de México, Mexico. [26]Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Mexico. [27]Instituto Oswaldo Cruz (FIOCRUZ), Rio de Janeiro, Brazil. [28]Paul-Ehrlich-Institut, Langen, Germany. [29]Laboratório Central de Saúde Pública Noel Nutels, Rio de Janeiro, Brazil. [30]Laboratório Central de Saúde Pública do Estado do Rio Grande do Norte, Natal, Brazil. [31]Universidade Potiguar do Rio Grande do Norte, Natal, Brazil. [32]Faculdade Natalense de Ensino e Cultura, Rio Grande do Norte, Natal, Brazil. [33]Laboratório Central de Saúde Pública do Estado da Paraíba, João Pessoa, Brazil. [34]Fundação Oswaldo Cruz (FIOCRUZ), Recife, Pernambuco, Brazil. [35]Department of Microbiology, Immunology & Pathology, Colorado State University, Fort Collins, Colorado 80523, USA. [36]Center for Vaccine Research, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [37]Section Clinical Tropical Medicine, Department for Infectious Diseases, Heidelberg University Hospital, Heidelberg, Germany. [38]Laboratório Central de Saúde Pública do Estado de Alagoas, Maceió, Brazil. [39]Universidade Estadual de Feira de Santana, Feira de Santana, Bahia, Brazil. [40]Secretaria de Saúde de Feira de Santana, Feira de Santana, Bahia, Brazil. [41]Universidade Federal do Amazonas, Manaus, Brazil. [42]Hospital São Francisco, Ribeirão Preto, Brazil. [43]Universidade Federal do Tocantins, Palmas, Brazil. [44]University of Sydney, Sydney, Australia. [45]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, UK. [46]Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA. [47]Department of Pathology, University of Texas Medical Branch, Galveston, Texas 77555, USA. [48]Metabiota, San Francisco, California 94104, USA.
*These authors contributed equally to this work.
§These authors jointly supervised this work.

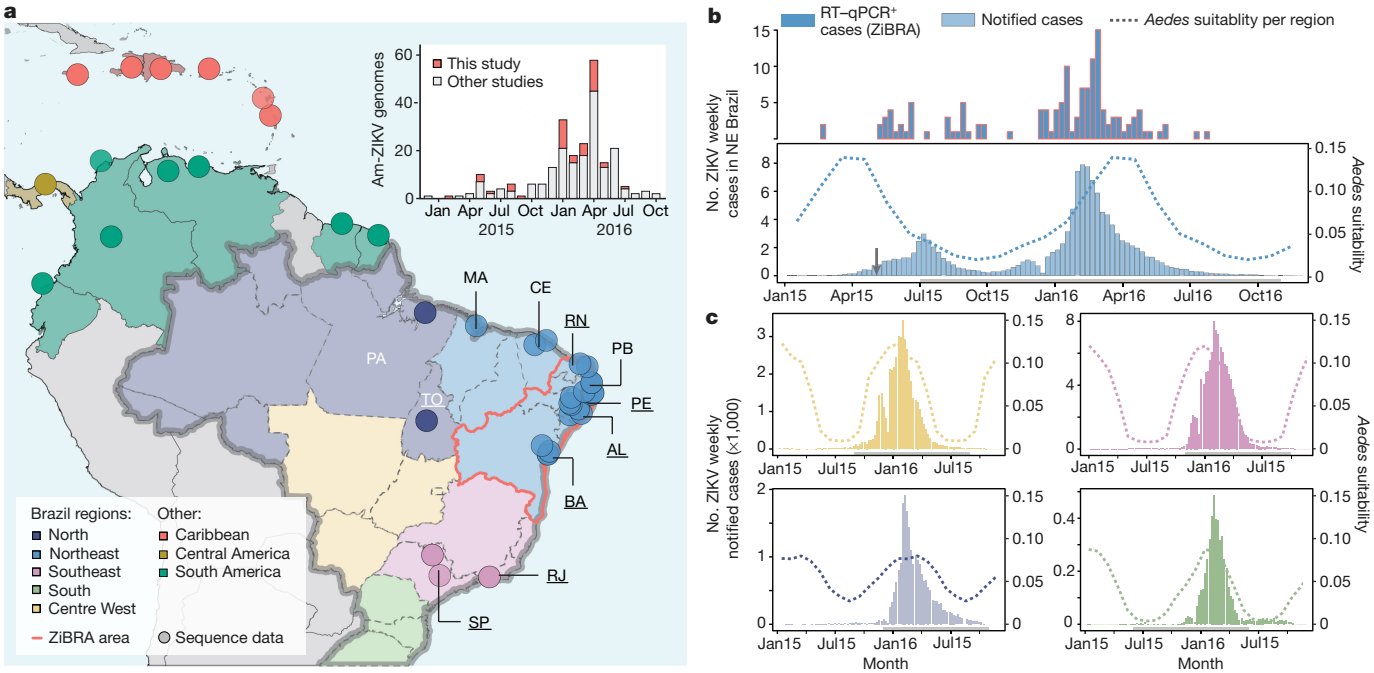


Figure 1 | Geographic and temporal distribution of ZIKV in Brazil. **a**, Sampling locations of genome sequences from Brazil and the Americas. Federal states in Brazil are coloured according to five geographic regions (lower inset). A red line surrounds the states surveyed by the ZiBRA mobile laboratory in 2016. State codes: AL, Alagoas; BA, Bahia; CE, Ceará; MA, Maranhão; PA, Pará; PB, Paraíba; PE, Pernambuco; RJ, Rio de Janeiro; RN, Rio Grande do Norte; SP, São Paulo; TO, Tocantins. Underlined states represent those from which sequences in this study were generated (upper inset). Publicly available sequences were also collated from non-underlined states. **b**, Confirmed and notified ZIKV cases in northeast (NE) Brazil. Upper panel shows the temporal distribution of RT–qPCR$^+$ cases detected during ZiBRA fieldwork. Only samples with known collection dates are included ($n = 138$ out of 181 confirmed cases). Lower panel shows notified ZIKV cases in northeast Brazil between 1 January 2015 and 19 November 2016 ($n = 122,779$). The dashed line represents the average climatic vector suitability score for northeast Brazil (see Methods). The vertical arrow indicates date of ZIKV confirmation in northeast Brazil and the Americas[1]. **c**, Notified ZIKV cases in the centre-west ($n = 44,825$), southeast ($n = 112,689$), north ($n = 22,373$), and south ($n = 4,944$) regions of Brazil (clockwise from top left). The dashed lines represent the average climatic vector suitability score for each region.

Ministry of Health and other institutions (see Acknowledgements), the ZiBRA laboratory screened 1,330 samples (almost exclusively serum or blood) from patients in 82 municipalities across 5 federal states (Fig. 1, Extended Data Table 1a). Samples provided by the public health laboratories of each state (LACEN) and the Fundação Oswaldo Cruz (FIOCRUZ) were screened for the presence of ZIKV by real-time quantitative PCR (RT–qPCR).

On average, ZIKV viraemia persists for 10 days after infection; symptoms develop after about 6 days and can last for 1–2 weeks[10]. In line with previous observations in Colombia[11], we found that RT–qPCR-positive samples from northeast Brazil were, on average, collected only 2 days after the onset of symptoms. The median RT–qPCR cycle threshold (C_t) value of positive samples was correspondingly high, at 36 (Extended Data Fig. 1a, b). For northeast Brazil, the time series of RT–qPCR$^+$ cases was positively correlated with the number of weekly notified cases (Pearson's $\rho = 0.62$; Fig. 1b).

The ability of the mosquito vector *Aedes aegypti* to transmit ZIKV is determined by ecological factors that affect adult survival, viral replication, and infective periods[12]. To investigate the receptivity of Brazilian regions to ZIKV transmission we used a measure of vector climatic suitability, derived from monthly temperature, relative humidity, and precipitation data[13]. Using linear regression we find that, for each Brazilian region, there is a strong association between estimated climatic suitability and weekly notified cases (Fig. 1b, c; adjusted $R^2 > 0.84$, $P < 0.001$; Extended Data Table 1b). Similar to previous findings from dengue virus outbreaks[14,15], notified ZIKV cases lag climatic suitability by about 4–6 weeks in all regions, except northeast Brazil, where no time lag is evident. Despite these associations, numbers of notified cases should be interpreted cautiously because co-circulating dengue and chikungunya viruses exhibit symptoms similar to ZIKV, and the Brazilian case reporting system has evolved through time (see Methods). We estimated basic reproductive numbers (R_0) for ZIKV in each Brazilian region from the weekly notified case data and found that R_0 was high in northeast Brazil ($R_0 \approx 3$ for both epidemic seasons; Extended Data Table 1c). Although our R_0 values are approximate, in part owing to spatial variation in transmission across the large regions analysed here, they are consistent with estimates from other approaches[16,17].

Encouraged by the utility of portable genomic technologies during the West African Ebola virus epidemic[18] we used our open protocol[19] to sequence ZIKV genomes directly from clinical material using MinION DNA sequencers. We were able to generate virus sequences within 48 h of the mobile laboratory's arrival at each LACEN. In pilot experiments using a cultured ZIKV reference strain[20] we recovered 98% of the virus genome (Extended Data Fig. 1c). However, owing to low viral copy numbers in clinical samples (Extended Data Fig. 1a), many sequences exhibited incomplete genome coverage and required additional sequencing efforts in static labs once fieldwork had been completed. Whereas average genome coverage was typically high for samples with lower C_t values (85% for $C_t < 33$; Fig. 2a, Extended Data Table 2), samples with higher C_t values had variable coverage (mean 72% for $C_t \geq 33$; Fig. 2a). Unsequenced genome regions were non-randomly distributed (Fig. 2b), suggesting that the efficiency of PCR amplification varied among primer pair combinations. We generated 36 near-complete or partial genomes from the northeast, southeast and northern regions of Brazil, supplemented by nine sequences from samples from Rio de Janeiro municipality. To further reconstruct Zika virus transmission in the Americas, we include five new complete ZIKV genomes from

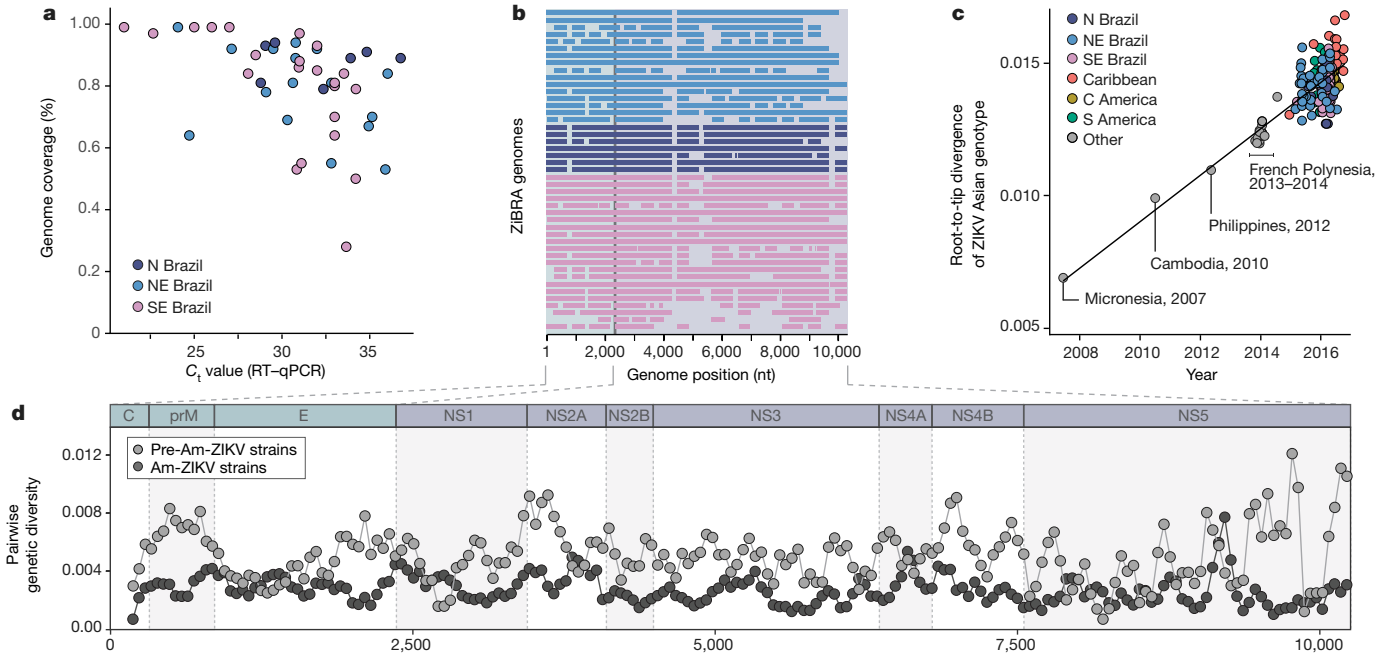


Figure 2 | Zika virus genetic diversity and sequencing statistics. a, The percentage of ZIKV genome sequenced plotted against RT–qPCR C_t value for each sample ($n = 45$). Each circle represents a sequence recovered from an infected individual in Brazil and is coloured by sampling location. N Brazil, North Brazil; NE Brazil, northeast Brazil; SE Brazil, southeast Brazil. **b**, Illustration of sequencing coverage across the ZIKV genome for the ZiBRA sequences, including data generated by both mobile and static laboratories. **c**, Regression of sequence sampling dates against root-to-tip genetic distances in a maximum likelihood phylogeny of the Asian-ZIKV lineage ($n = 254$). Extended Data Fig. 2b contains a comparable analysis that also includes P6-740 (the oldest Asian-ZIKV strain collected in 1966) ($n = 255$). **d**, Average pairwise genetic diversity of the PreAm-ZIKV strains ($n = 19$, grey line) and of the Am-ZIKV lineage ($n = 235$, black line), calculated using a sliding window of 300 nucleotides with a step size of 50 nucleotides.

Colombia and four from Mexico. In addition, we append to our dataset 115 publicly available sequences and 85 additional genomes from ref. 21. The final dataset comprised 254 ZIKV sequences, 241 of which were sampled in the Americas (see Methods).

The American ZIKV epidemic comprises a single founder lineage[4,22,23] (hereafter termed Am-ZIKV) derived from Asian genotype viruses (hereafter termed PreAm-ZIKV) from southeast Asia and the Pacific[4]. A sliding window analysis of pairwise genetic diversity along the ZIKV genome shows that the diversity of PreAm-ZIKV strains is on average about two-fold greater than that of Am-ZIKV viruses (Fig. 2d), reflecting a longer period of ZIKV circulation in Asia and the Pacific than in the Americas. The genetic diversity of Am-ZIKV strains will increase in the future and updated diagnostic assays are recommended to guarantee RT–qPCR sensitivity[24].

It has been suggested that recent ZIKV epidemics may be linked causally to a higher apparent evolutionary rate for the Asian genotype than the African genotype[25,26]. However, such comparisons are confounded by an inverse relationship between the timescale of observation and estimated evolutionary rates[27]. Regression of sequence sampling dates against root-to-tip genetic distances indicates that molecular clock models can be applied reliably to the Asian ZIKV lineage (Fig. 2c, Extended Data Figs 2, 3). We estimate the whole-genome evolutionary rate of Asian ZIKV to be 1.12×10^{-3} substitutions per site per year (95% Bayesian credible interval (BCI) $0.97–1.27 \times 10^{-3}$), consistent with other estimates for this lineage[4,26]. We found no significant differences in evolutionary rates among ZIKV genome regions (Extended Data Table 3a). The estimated ratio of divergence at non-synonymous and synonymous sites (d_N/d_S) of the Am-ZIKV lineage is low (0.11, 95% confidence interval 0.10–0.13), as observed for other vector-borne flaviviruses[28], but is higher than that of PreAm-ZIKV viruses (0.061, 0.047–0.077), probably owing to the raised probability of observing slightly deleterious changes in short-term datasets, as observed during previous epidemics[29].

We used two phylogeographic approaches with different assumptions[30,31] to reconstruct the origins and spread of ZIKV in Brazil and the Americas. We dated the common ancestor of ZIKV in the Americas (node B, Fig. 3) to Jan 2014 (95% BCI October 2013–April 2014; Extended Data Tables 3b, c), in line with previous estimates[4,26]. We find evidence that northeast Brazil played a central role in the establishment and dissemination of Am-ZIKV. Although northeast Brazil is the most probable location of node B (location posterior support 0.83, Fig. 3), the current data do not allow us to exclude the hypothesis that node B was in the Caribbean (Fig. 3 dashed branches) owing to the presence of two sequences from Haiti in one of its descendant lineages. More importantly, most Am-ZIKV sequences descend from a radiation of lineages (node C and its immediate descendants; Fig. 3) dated to late February 2014 (95% BCIs of node C, November 2013–May 2014). Node C is more strongly inferred to have existed in northeast Brazil (location posterior support 0.99, Fig. 3). All 20 replicate analyses performed on subsampled datasets place node C in Brazil, and 14 of them place node C in northeast Brazil (Extended Data Fig. 4). Consequently, we conclude that node C reflects the crucial turning point in the emergence of ZIKV in the Americas. If further data show that node B did exist in Haiti, then it is likely that Haiti acted as an intermediate 'stepping stone' for the arrival and establishment of Am-ZIKV in Brazil, from where the virus subsequently spread to other regions. This perspective is consistent with the lower population size of Haiti compared to Brazil. We infer that node C was present in northeast Brazil several months before three notable events, each of which also occurred in northeast Brazil: (i) the retrospective identification of a cluster of suspected but unconfirmed ZIKV cases in December 2014[1]; (ii) the collection of the oldest ZIKV genome sequence from Brazil, reported here, sampled in February 2015; and (iii) confirmation of cases of ZIKV transmission in northeast Brazil in March 2015[32,33].

Our results further indicate that viruses from northeast Brazil were important for the continental spread of ZIKV. Within Brazil, we find

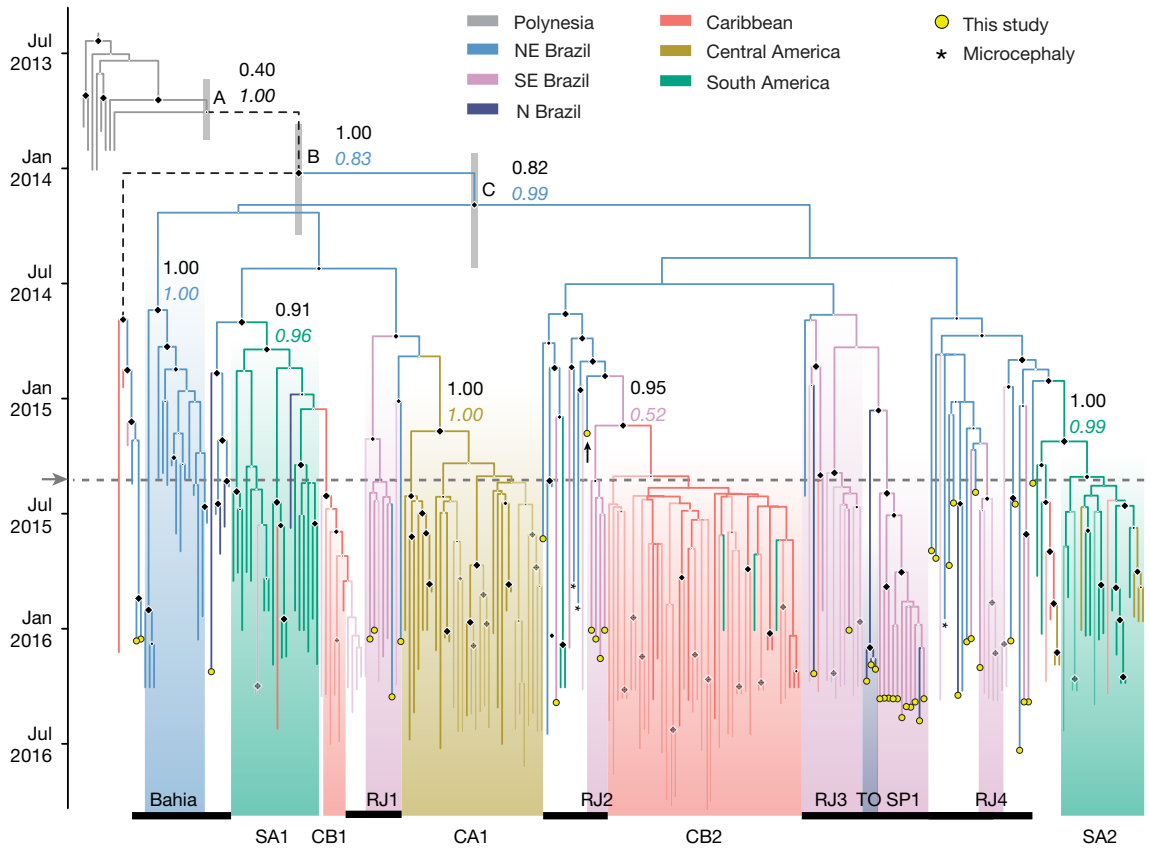


Figure 3 | Phylogeography of ZIKV in the Americas. Maximum clade credibility phylogeny, estimated from complete and partial Am-ZIKV genomes using a molecular clock phylogeographic approach (see Methods). Terminal branches with yellow circles indicate sequences reported in this study. Terminal branches with no circles and reduced opacity are those reported in ref. 21. Thin vertical grey boxes indicate statistical uncertainty of estimated dates of nodes A, B and C (Extended Data Table 3c). Branch colours indicate the most probable ancestral lineage locations. Diamonds at internal nodes are sized in proportion to clade posterior probabilities. For selected nodes, coloured numbers show the posterior probabilities of ancestral locations and numbers in black are clade posterior probabilities. Asterisks indicate the three available genomes from microcephaly cases. A black arrow indicates the oldest Brazilian ZIKV sequence. The grey arrow and dotted line denote when ZIKV was first confirmed in the Americas[1]. Nodes A and B are equivalent to the nodes named identically in ref. 4. Text labels along the bottom of the figure denote clades of sequences from regions outside northeast Brazil. RJ1–RJ4 are clades from Rio de Janeiro state, TO from Tocantins, and SP1 from São Paulo state. Clades from outside Brazil are denoted CB1 and CB2 (Caribbean), SA1 and SA2 (South America excluding Brazil), and CA1 (Central America). Black horizontal lines along the bottom of the figure denote sequences from Brazil.

instances of virus lineage movement from northeast to southeast Brazil; most of these events are dated to the second half of 2014 and led to onwards transmission in Rio de Janeiro (RJ1–RJ4; Fig. 3) and São Paulo states (SP1; Fig. 3). We infer that ZIKV lineages disseminated from northeast Brazil to elsewhere in Central America, the Caribbean, and South America. Most Am-ZIKV strains sampled outside Brazil fall into four well-supported phylogenetic groups (Fig. 3); three (SA1/CB1, CA1 and SA2) are inferred to have been exported from northeast Brazil between July 2014 and April 2015, whereas the Caribbean clade CB2 appears to have originated from southeast Brazil around March 2015 (Figs 3, 4). Each viral lineage export occurred during a period of climatic suitability for vector transmission in the

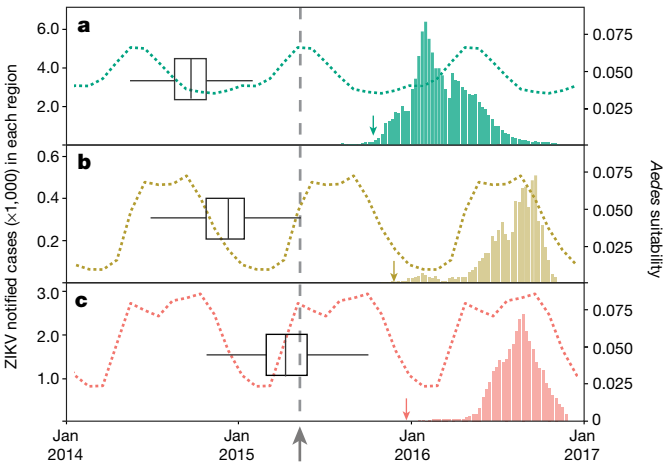


Figure 4 | Establishment of Am-ZIKV in the Americas. The earliest inferred dates of lineage export to non-Brazilian regions, represented by box and whisker plots. Each plot corresponds to the earliest movement between a pair of locations with well-supported virus lineage migration. The first exports to South America outside Brazil (SA1 in Fig. 3), to Central America (CA1) and to the Caribbean (CB1) are shown in **a**–**c**, respectively. Box and whisker plots were generated in ggplot2, with boxes representing the median and interquartile ranges of the estimated date of earliest movement. In each of **a**–**c**, dashed lines show the estimated climatic vector suitability score for each recipient region, averaged across the countries for which sequence data are available (see Methods). In each of **a**–**c**, the bar plots show available notified ZIKV case data (plots adapted from PAHO) for the countries with the earliest confirmed cases (Colombia in **a**, Mexico in **b**, and Puerto Rico in **c**, see Methods). Coloured arrows indicate the earliest confirmation of ZIKV autochthonous cases in each non-Brazilian region. The vertical dashed line represents the date of ZIKV confirmation in the Americas.

recipient location (Fig. 4). For the earliest exports to Central America (CA1) and South America (SA1), there is an estimated 11–12-month gap between the date of export and the date of ZIKV detection in the recipient location, suggesting a complete season of undetected transmission. These periods of cryptic transmission are relevant to studies of spatiotemporal trends in reported microcephaly, because they help to define the appropriate timeframe for baseline (pre-ZIKV) microcephaly in each region.

Large-scale surveillance of ZIKV is challenging because many cases may be asymptomatic, and ZIKV co-circulates in some regions with other arthropod-borne viruses that have overlapping symptoms (for example, dengue, chikungunya, Mayaro, and Oropouche viruses). However combining virus genomic and epidemiological data can generate insights into vector-borne virus transmission. A system of continuous and structured virus sequencing in Brazil, integrated with surveillance data, could provide timely information to inform effective responses against Zika and other viruses, including the recently re-emerged yellow fever virus[34].

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 8 February; accepted 2 May 2017.
Published online 24 May 2017.

1. Kindhauser, M. K., Allen, T., Frank, V., Santhana, R. S. & Dye, C. Zika: the origin and spread of a mosquito-borne virus. *Bull. World Health Organ.* **94,** 675–686C (2016).
2. Ministério da Saúde. Boletins Epidemiológicos—Secretaria de Vigilância em Saúde http://portalsaude.saude.gov.br/index.php/o-ministerio/principal/secretarias/svs/boletim-epidemiologico (2017).
3. WHO. Situation Report—Zika virus, microcephaly, Guillain–Barré syndrome (10 March 2017) http://apps.who.int/iris/bitstream/10665/254714/1/zikasitrep10Mar17-eng.pdf?ua=1 (2017).
4. Faria, N. R. *et al.* Zika virus in the Americas: early epidemiological and genetic findings. *Science* **352,** 345–349 (2016).
5. Alex Perkins, T., Siraj, A. S., Ruktanonchai, C. W., Kraemer, M. U. & Tatem, A. J. Model-based projections of Zika virus infections in childbearing women in the Americas. *Nat. Microbiol*. **1,** 16126 (2016).
6. Lessler, J. *et al.* Assessing the global threat from Zika virus. *Science* **353,** aaf8160 (2016).
7. Vasconcelos, P. F. & Calisher, C. H. Emergence of human arboviral diseases in the Americas, 2000–2016. *Vector Borne Zoonotic Dis*. **16,** 295–301 (2016).
8. Vogel, G. One year later, Zika scientists prepare for a long war. *Science* **354,** 1088–1089 (2016).
9. Bogoch, I. I. *et al.* Anticipating the international spread of Zika virus from Brazil. *Lancet* **387,** 335–336 (2016).
10. 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).
11. Pacheco, O. *et al.* Zika virus disease in Colombia—preliminary report. *New Engl. J. Med.* http://dx.doi.org/10.1056/NEJMoa1604037 (2016).
12. Liu-Helmersson, J., Stenlund, H., Wilder-Smith, A. & Rocklöv, J. Vectorial capacity of *Aedes aegypti*: effects of temperature and implications for global dengue epidemic potential. *PLoS One* **9,** e89783 (2014).
13. 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.* **16,** 1237–1245 (2016).
14. Cuong, H. Q. *et al.* Quantifying the emergence of dengue in Hanoi, Vietnam: 1998–2009. *PLoS Negl. Trop. Dis.* **5,** e1322 (2011).
15. Gharbi, M. *et al.* Time series analysis of dengue incidence in Guadeloupe, French West Indies: forecasting models using climate variables as predictors. *BMC Infect. Dis.* **11,** 166 (2011).
16. Caminade, C. *et al.* Global risk model for vector-borne transmission of Zika virus reveals the role of El Niño 2015. *Proc. Natl Acad. Sci. USA* **114,** 119–124 (2017).
17. Rocklöv, J. *et al.* Assessing seasonal risks for the introduction and mosquito-borne spread of Zika virus in Europe. *EBioMedicine* **9,** 250–256 (2016).
18. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530,** 228–232 (2016).
19. Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protocols* http://dx.doi.org/10.1038/nprot.2017.066 (2017).
20. Trösemeier, J. H. *et al.* Genome sequence of a candidate World Health Organization reference strain of Zika virus for nucleic acid testing. *Genome Announc*. **4,** e00917–16 (2016).
21. Metsky, H. C. *et al.* Zika virus evolution and spread in the Americas. *Nature* http://dx.doi.org/10.1038/nature22402 (2017).
22. Giovanetti, M. *et al.* Zika virus complete genome from Salvador, Bahia, Brazil. *Infect. Genet. Evol.* **41,** 142–145 (2016).
23. Naccache, S. N. *et al.* Distinct Zika virus lineage in Salvador, Bahia, Brazil. *Emerg. Infect. Dis.* **22,** 1788–1792 (2016).
24. Corman, V. M. *et al.* Assay optimization for molecular detection of Zika virus. *Bull. World Health Organ.* **94,** 880–892 (2016).
25. Liu, H. *et al.* From discovery to outbreak: the genetic evolution of the emerging Zika virus. *Emerg. Microbes Infect*. **5,** e111 (2016).
26. Pettersson, J. H. O. *et al.* How did Zika virus emerge in the Pacific Islands and Latin America? *MBio* **7,** e01239–16 (2016).
27. Holmes, E. C., Dudas, G., Rambaut, A. & Andersen, K. G. The evolution of Ebola virus: Insights from the 2013–2016 epidemic. *Nature* **538,** 193–200 (2016).
28. Holmes, E. C. Patterns of intra- and interhost nonsynonymous variation reveal strong purifying selection in dengue virus. *J. Virol.* **77,** 11296–11298 (2003).
29. Park, D. J. *et al.* Ebola virus epidemiology, transmission, and evolution during seven months in Sierra Leone. *Cell* **161,** 1516–1526 (2015).
30. De Maio, N., Wu, C. H., O'Reilly, K. M. & Wilson, D. New routes to phylogeography: a Bayesian structured coalescent approximation. *PLoS Genet.* **11,** e1005421 (2015).
31. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its roots. *PLOS Comput. Biol.* **5,** e1000520 (2009).
32. Campos, G. S., Bandeira, A. C. & Sardi, S. I. Zika virus outbreak, Bahia, Brazil. *Emerg. Infect. Dis.* **21,** 1885–1886 (2015).
33. Zanluca, C. *et al.* First report of autochthonous transmission of Zika virus in Brazil. *Mem. Inst. Oswaldo Cruz* **110,** 569–572 (2015).
34. Paules, C. I. & Fauci, A. S. Yellow fever—once again on the radar screen in the Americas. *N. Engl. J. Med.* **376,** 1397–1399 (2017).

Acknowledgements We thank Fundação Oswaldo Cruz in Bahia and Pernambuco states, University of São Paulo, Instituto Evandro Chagas, and the Brazilian Zika virus surveillance network for their essential contributions. We thank the following for giving us permission to use their unpublished genomes available on GenBank: R. Lanciotti, J. Lednicky, A. Enfissi, F. Baldanti, R. Shabman, B. Picket, R. Schinazi, M. Bonaldo, M. Gale, M. Capobianchi and C. Concetta, M. Leguia, J. Alberto Diaz, E. Sevilla-Reyes, A. Franz, M. Garcia-Blanco and M. J. van Hemert. We thank P. Fernando da Costa Vasconcelos, S. Guerreiro Rodrigues, J. Cardoso, J. Vasconcelos, J. Vianez Jr, J. Gil Melgaço, J. Blumel, M. C. Brito Lobato, L. Nunes Fava, C. Ayres, L. Abade and F. Campos. L.C.J.A. thanks QIAGEN for reagents and equipment and M.R.T.N. thanks FERPEL for consumables. We thank Oxford Nanopore for technical support, particularly R. Dokos, Z. McDougall, S. Cowan, G. Sanghera, and O. Hartwell. This work was supported by an MRC/Wellcome Trust/Newton Fund Zika Rapid Response grant (MC_PC_15100/ZK/16-078) and by the USAID Emerging Pandemic Threats Program-2 PREDICT-2 (Cooperative Agreement AID-OAA-A-14-00102). N.J.L. is supported by an MRC Bioinformatics Fellowship. N.R.F. is funded by a Sir Henry Dale Fellowship (grant 204311/Z/16/Z). CNPq contributed to trip expenses (grant 457480/2014-9). A.C.d.C. was supported by FAPESP #2012/03417-7 and M.R.T.N. by CNPq grant no. 302584/2015-3. A.B. and T.B. were supported by NIH award R35 GM119774. A.B. is supported by the NSF Graduate Research Fellowship Program (grant DGE-1256082). T.B. is a Pew Biomedical Scholar. C.Y.C. is partially supported by NIH grant R01 HL105704 and an award from Abbott Laboratories, Inc. E.C.H. is supported by a National Health and Medical Research Council Australia Fellowship (GNT1037231). C.-H.W. is supported by the MRC and CRUK (ANR00310) and by the Wellcome Trust and Royal Society (grant 101237/Z/13/Z). S.C.H. is supported by the Wellcome Trust. This research received funding from the ERC under grant agreements 614725-PATHPHYLODYN and 278433-PREDEMICS, and from EU Horizon 2020 under agreements 643476-COMPARE and 734548-ZIKAlliance. T.J. and E.T.M. acknowledge funding from IDAMS, DENFREE, DengueTools, and PPSUS-FACEPE (project APQ-0302-4.01/13). R.F.F. received funding from FACEPE (APQ-0044.2.11/16 and APQ-0055.2.11/16) and from CNPq (439975/2016-6). S.A.B. was supported by the Sicherheit von Blut und Geweben hinsichtlich der Abwesenheit von Zikaviren from the German Ministry of Health.

Author Contributions N.R.F., L.C.J.A., M.R.T.N., E.C.S., N.J.L. and O.G.P. designed the study. N.R.F., J.Q., N.J.L., I.M.C., J.G.d.J., M.G., S.C.H., A.B., A.C.d.C., L.C.F., S.P.S., T.B., P.S.L., B.L.S.N., H.A.O.M., M.R.T.N. and L.C.J.A. undertook fieldwork and experiments. N.R.F., J.T., C.-H.W., O.G.P., J.R. and L.d.P. performed genetic analyses. N.R.F., M.U.G.K., O.G.P. and S.C. performed epidemiological analyses. N.R.F., J.Q., M.U.G.K., N.J.L. and O.G.P. wrote the manuscript. E.C.H., A.R., T.B., M.R.T.N., E.C.S. and L.C.J.A. edited the manuscript. All other authors were critical for coordination, collection, processing, sequencing and bioinformatics of samples. All authors read and approved the contents of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. 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 L.C.J.A. (lalcan@bahia.fiocruz.br), E.C.S. (sabinoec@usp.br), N.J.L. (n.j.loman@bham.ac.uk) and O.G.P. (oliver.pybus@zoo.ox.ac.uk).

Reviewer Information *Nature* thanks K. St George, A. Wilder-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.

Sample collection. Between 1 and 18 June 2016, 1,330 samples from cases notified as ZIKV infected were tested for ZIKV infection in the northeast region of Brazil. During this period, four of the five laboratories in the region visited by the ZiBRA project were in the process of implementing molecular diagnostics for ZIKV. The ZiBRA team spent 2–3 days in each state central public health laboratory (LACEN). The samples analysed had been previously collected from patients who had attended a municipal or state public health facility, presenting maculopapular rash and at least two of the following symptoms: fever, conjunctivitis, polyarthralgia, or periarticular oedema. The majority of samples were linked to a digital record that collated epidemiological and clinical data: date of sample collection, location of residence, demographic characteristics, and date of onset of clinical symptoms (when available).

The ZiBRA project was supported by the Brazilian Ministry of Health (MoH) as part of the emergency public health response to Zika. Samples had been previously obtained for routine diagnostic purposes from persons visiting local clinics by the Brazilian National Health Surveillance network as part of Zika virus surveillance activities. In these cases, we used samples without informed consent with the approval of the Brazilian Ministry of Health. Specifically, residual anonymized clinical diagnostic samples, with no or minimal risk to patients, were provided for research and surveillance purposes within the terms of Resolution 510/2016 of CONEP (Comissão Nacional de Ética em Pesquisa, Ministério da Saúde; National Ethical Committee for Research, Ministry of Health). For samples obtained from patients engaged in longitudinal studies of Zika virus in São Paulo and Tocantins states, informed consent was obtained (IRB CAAE 53153916.7.0000.0065). Samples from patients followed in Salvador and Feira de Santana were analysed under institutional approval from CPqGM/FioCruz/BA (1.184.454). Urine and plasma samples from Rio de Janeiro were obtained from patients at the Fiocruz Viral Hepatitis Ambulatory (Oswaldo Cruz Institute, Rio de Janeiro, Brazil) with Institutional Review Board approval (IRB142/01) from the Oswaldo Cruz Institute. RNA was extracted at the Paul-Ehrlich-Institut and sequenced at the University of Birmingham, UK.

Nucleic acid isolation and RT–qPCR. Serum, blood and urine samples were obtained from patients 0–228 days after their first symptoms (Extended Data Table 1a). Viral RNA was isolated from 200-μl Zika-suspected samples using the NucliSENS easyMag system (BioMerieux, Basingstoke, UK) (Ribeirão Preto samples), the ExiPrep Dx Viral RNA Kit (BIONEER, Republic of Korea) (Rio de Janeiro samples) or the QIAamp Viral RNA Mini kit (QIAGEN, Hilden, Germany) (all other samples) according to the manufacturer's instructions. C_t values were determined for all samples by probe-based RT–qPCR against the *prM* target (using 5′ FAM as the probe reporter dye) as previously described[35]. RT–qPCR assays were performed using the QuantiNova Probe RT–qPCR Kit (20-μl reaction volume; QIAGEN) with amplification in the Rotor-Gene Q (QIAGEN) following the manufacturer's protocol. Primers and probes were synthesized by Integrated DNA Technologies (Leuven, Belgium). The following reaction conditions were used: reverse transcription (50 °C, 10 min), reverse transcriptase inactivation and DNA polymerase activation (95 °C, 20 s), followed by 40 cycles of DNA denaturation (95 °C, 10 s) and annealing–extension (60 °C, 40 s). Positive and negative controls were included in each batch; however, owing to the large number of samples tested in a short time, it was possible only to run each sample without replication.

Whole-genome sequencing. Sequencing was attempted on all positive samples obtained from northeast Brazil regardless of C_t value. All samples collected in Brazil that are reported in this study were sequenced with Oxford Nanopore MinION. Sequencing statistics can be found in Extended Data Table 2. The protocol employed cDNA synthesis with random primers followed by gene specific multiplex PCR and is presented in detail in ref. 19. In brief, extracted RNA was converted to cDNA using the Protoscript II First Strand cDNA synthesis Kit (New England Biolabs, Hitchin, UK) and random hexamer priming. ZIKV genome amplification by multiplex PCR was attempted using the ZikaAsianV1 primer scheme and 40 cycles of PCR using Q5 High-Fidelity DNA polymerase (NEB) as described in ref. 19. PCR products were cleaned up using AmpureXP purification beads (Beckman Coulter, High Wycombe, UK) and quantified using fluorimetry with the Qubit dsDNA High Sensitivity assay on the Qubit 3.0 instrument (Life Technologies). PCR products for samples yielding sufficient material were barcoded and pooled in an equimolar fashion using the Native Barcoding Kit (Oxford Nanopore Technologies, Oxford, UK). Sequencing libraries were generated from the barcoded products using the Genomic DNA Sequencing Kit SQK-MAP007/SQK-LSK208 (Oxford Nanopore Technologies). Sequencing libraries were loaded onto a R9/R9.4 flow cell and data were collected for up to 48 h but generally less. As described[19], consensus genome sequences were produced by alignment of two-direction reads to a Zika virus reference genome (strain H/PF/2013, GenBank Accession number: KJ776791) followed by nanopore signal-level detection of single nucleotide variants. Only positions with $\geq 20\times$ genome coverage were used to produce consensus alleles. Regions with lower coverage, and those in primer-binding regions were masked with N characters. Validation of our sequencing approach on the MinION platform was undertaken by using the MinION platform to sequence a WHO reference strain of Zika virus that was also sequenced using the Illumina Miseq platform[20]; identical consensus sequences were recovered regardless of the MinION chemistry version employed (R7.3, R9 and R9.4) (Extended Data Fig. 1c).

Collation of genome-wide datasets. Our complete and partial genome sequences were appended to a global dataset of all available published ZIKV genome sequences (up until 1 March 2017) using an in-house script that retrieves updated GenBank sequences on a daily basis. In addition to the genomes generated from samples collected in northeast Brazil during ZiBRA fieldwork, samples were sent directly to the University of São Paulo and elsewhere for sequencing. Thirteen genomes from Ribeirão Preto, São Paulo state (SP; southeast Brazil) and seven genomes from Tocantins (TO; north Brazil) were sequenced at the University of São Paulo. Nine genomes from Rio de Janeiro (RJ; southeast Brazil) were sequenced in Birmingham, UK, and added to our dataset. All these genomes were generated using the same primer scheme as the ZiBRA samples collected in northeast Brazil[18]. In addition to these 45 sequences from Brazil, we further included in analysis 9 genomes from ZIKV strains sampled outside Brazil in order to contextualise the genetic diversity of Brazilian ZIKV, giving rise to a final dataset of 54 sequences. Specifically, we included five genomes from samples collected in Colombia and four new genomes from Mexico, which were generated using the protocols described in refs 36 and 23, respectively.

GenBank sequences belonging to the African genotype of ZIKV were identified using the Arboviral genotyping tool (http://bioafrica2.mrc.ac.za/rega-genotype/typingtool/aedesviruses) and excluded from subsequent analyses, as our focus of study was the Asian genotype of ZIKV, and the Am-ZIKV lineage in particular. To assess the robustness of molecular clock dating estimates to the inclusion of older sequences, analyses were performed both with and without the P6-740 strain, the oldest known strain of the ZIKV-Asian genotype (sampled in 1966 in Malaysia). Our final alignment comprised the sequences reported in this study ($n = 54$) plus publicly available ZIKV-Asian genotype sequences, as of 1 March 2017 ($n = 115$). We also included in our analysis 85 additional genomes from ref. 21. The dataset used for analysis therefore included sequences from 254 Zika virus isolates, 241 of which were from the Americas. Unpublished but publicly available genomes were included in our analysis only if we had written permission from those who generated the data (see Acknowledgements).

Maximum likelihood analysis and recombination screening. Preliminary maximum likelihood (ML) trees were estimated with ExaML version 3 (ref. 37) using a per-site rate category model and a gamma distribution of among site rate variation. For the final analyses, ML trees were estimated using PhyML[38] under a GTR nucleotide substitution model[39], with a gamma distribution of among site rate variation, as selected by jModelltest version 2 (ref. 40). Branch support was inferred using 100 bootstrap replicates[37]. Final ML trees were estimated with NNI and SPR heuristic tree search algorithms; equilibrium nucleotide frequencies and substitution model parameters were estimated using ML[38] (see Extended Data Fig. 3).

Recombination may impact evolutionary estimates[41] and has been shown to be present in the ZIKV-African genotype[42]. In addition to restricting our analysis to the Asian genotype of ZIKV, we employed the 12 recombination detection methods available in RDP version 4 (ref. 42) and the Phi-test approach[44] available in SplitsTree[45] to further search for evidence of recombination in the ZIKV-Asian lineage. No evidence of recombination was found.

Analysis of the temporal molecular evolutionary signal in our ZIKV alignments was conducted using TempEst[46]. In brief, collection dates in the format yyyy-mm-dd (ISO 8601 standard) were regressed against root-to-tip genetic distances obtained from the ML phylogeny. When precise sampling dates were not available, a precision of 1 month or 1 year in the collection dates was taken into account.

To compare the pairwise genetic diversity of PreAm-ZIKV strains from Asia and the Pacific with Am-ZIKV viruses from the Americas, we used a sliding window approach with 300-nt wide windows and a step size of 50 nt. Sequence gaps were ignored; hence the average pairwise difference per window was obtained by dividing the total pairwise nucleotide differences by the total number of pairwise comparisons.

Molecular clock phylogenetics and gene-specific d_N/d_S estimation. To estimate Bayesian molecular clock phylogenies, analyses were run in duplicate using BEAST version 1.8.4 (ref. 47) for 30 million MCMC steps, sampling parameters and trees every 3,000 steps. We used a model selection procedure using both path-sampling and stepping-stone models[48] to estimate the most appropriate combination of

molecular clock and coalescent models for Bayesian phylogenetic analysis. The best fitting combination was a Bayesian skyline tree prior and a relaxed molecular clock model, with log-normally distributed variation in rates among branches (Extended Data Table 3b). A non-informative continuous time Markov chain reference prior[49] on the molecular clock rate was used. Convergence of MCMC chains was checked with Tracer version 1.6. After removal of burn-in, posterior tree distributions were combined and subsampled to generate an empirical distribution of 1,500 molecular clock trees.

To estimate rates of evolution per gene we partitioned the alignment into 10 genes (three structural genes *C*, *prM*, *E*, and seven non-structural genes *NS1*, *NS2A*, *NS2B*, *NS3*, *NS4A2K*, *NS4B* and *NS5*) and employed a SDR06 substitution model[50] and a strict molecular clock model, using an empirical distribution of molecular clock phylogenies. To estimate the ratio of nonsynonymous to synonymous substitutions per site (d_N/d_S) for the PreAm-ZIKV and the Am-ZIKV lineages, we used the single likelihood ancestor counting (SLAC) method[51] implemented in HyPhy[52]. This method was applied to two distinct codon-based alignments and their corresponding ML trees which comprised the PreAm-ZIKV and Am-ZIKV sequences, respectively.

Phylogeographic analysis. We investigated virus lineage movements using our empirical distribution of phylogenetic trees and the sampling location of each ZIKV sequence. The sampling location of sequences collected from returning travellers was set to the travel destination in the Americas where infection likely occurred. We discretised sequence sampling locations in Brazil into the geographic regions defined in the main text. The number of sequences per region available for analysis was 10 for north Brazil, 41 for northeast Brazil and 54 for southeast Brazil. No viral genetic data were available for the centre-west or south Brazilian regions. We similarly discretized the locations of ZIKV sequences sampled outside Brazil. These were grouped according to the United Nations M49 coding classification of macro-geographical regions. Our analysis included 53 sequences from the Caribbean, 38 from Central America, 17 from Polynesia, 37 from South America (excluding Brazil), 3 from Southeast Asia and 1 from Micronesia. To account for the possibility of sampling bias arising from a larger number of sequences from particular locations, we performed all phylogeographic analyses using (i) the full dataset ($n = 254$) and (ii) ten jackknife resampled datasets ($n = 74$) in which taxa from each location (except for Southeast Asia and Micronesia) were randomly sub-sampled to 10 sequences (the number of sequences available for north Brazil).

Phylogeographic reconstructions were conducted using two approaches; (i) using the asymmetric[53] discrete trait evolution models implemented in BEAST version 1.8.4 (ref. 47) and (ii) using the Bayesian structured coalescent approximation (BASTA)[30] implemented in BEAST2 version 2. The latter has been suggested to be less sensitive to sampling biases[54]. For both approaches, maximum clade credibility trees were summarized from the MCMC samples using TreeAnnotator after discarding 10% as burn-in. The posterior estimates of the location of nodes A, B and C (depicted in Fig. 3) from these two analytical approaches (applied to both the complete and jackknifed datasets) can be found in Extended Data Fig. 4.

For the discrete trait evolution approach, we counted the expected number of transitions among each pair of locations (net migration) using the robust counting approach[55,56] available in BEAST version 1.8.4 (ref. 47). We then used those inferred transitions to identify the earliest estimated ZIKV introductions into new regions. These viral lineage movement events were statistically supported (with Bayes factors >3) using the BSSVS (Bayesian stochastic search variable selection) approach implemented in BEAST version 1.8.4 (ref. 31). Box plots for node ages were generated using the ggplot2 (ref. 57) package in R software[58]. Case counts for shown in Fig. 4 were obtained from Pan American Health Organization epidemiological reports for Colombia, Mexico and Puerto Rico[62–64].

Epidemiological analysis. Weekly suspected ZIKV data per Brazilian region were obtained from the Brazilian Ministry of Health (MoH). Cases were defined as suspected ZIKV infection when patients presented maculopapular rash and at least two of the following symptoms: fever, conjunctivitis, polyarthralgia or periarticular oedema. Because notified suspected ZIKV cases are based on symptoms and not molecular diagnosis, it is possible that some notified cases represent other co-circulating viruses with related symptoms, such as dengue and chikungunya viruses. Furthermore, case reporting may have varied among regions and through time. Data from 2015 came from the pre-existing MoH sentinel surveillance system that comprised 150 reporting units throughout Brazil, which was eventually standardized in Feb 2016 in response to the ZIKV epidemic. We suggest that these limitations should be borne in mind when interpreting the ZIKV notified case data and we consider the R_0 values estimated here to be approximate. That said, our time series of RT–qPCR$^+$ ZIKV diagnoses from northeast Brazil qualitatively match the time series of notified ZIKV cases from the same region (Fig. 1b). To estimate the exponential growth rate of the ZIKV outbreak in Brazil, we fit a simple exponential growth rate model to each stage of the weekly number of suspected ZIKV cases from each region separately:

$$I_w = I_0 \exp(r_W \,.\, w) \tag{1}$$

where I_w is the number of cases in week w. As described in the main text, the Brazilian regions considered here were northeast Brazil, north Brazil, south Brazil, southeast Brazil, and centre-west Brazil. The time period over which exponential growth occured was determined by plotting the log of I_w and selecting the period of linearity (Extended Data Fig. 5). A linear model was then fitted to this period to estimate the weekly exponential growth rate r_w:

$$\ln(I_w) = \ln(I_0) + r_W \,.\, w \tag{2}$$

Let $g(\cdot)$ be the probability density distribution of the epidemic generation time (that is, the duration between the time of infection of a case and the mean time of infection of its secondary infections). The following formula can be used to derive the reproduction number R from the exponential growth rate r and density $g(\cdot)$ (ref. 59).

$$R = \frac{1}{\int_0^\infty \exp(-r \,.\, t) g(t) dt} \tag{3}$$

In our baseline analysis, following ref. 60, we assume that the ZIKV generation time is gamma-distributed with a mean of 20.0 days and a standard deviation (s.d.) of 7.4 days. In a sensitivity analysis, we also explored scenarios with shorter mean generation times (10.0 and 15.0 days) but unchanged coefficient of variation s.d./mean = 7.4/20 = 0.37 (Extended Data Table 1c).

Association between *Aedes aegypti* climatic suitability and ZIKV notified cases. To account for seasonal variation in the geographical distribution of the ZIKV vector *A. aegypti* in Brazil we fitted high-resolution maps[61] to monthly covariate data. Covariate data included time-varying variables, such as temperature persistence suitability, relative humidity, and precipitation, as well as static covariates such as urban versus rural land use. Maps were produced at a 5-km × 5-km resolution for each calendar month and then aggregated to the level of the five Brazilian regions used in this study (Extended Data Fig. 6). For consistency, we rescaled monthly suitability values so that the sum of all monthly maps equalled the annual mean map[13].

We then assessed the correlation between monthly *A. aegypti* climatic suitability and the number of weekly ZIKV notified cases in each Brazilian region, to test how well vector suitability explains the variation in the number of ZIKV notified cases. To account for the correlation in each Brazilian region we fit a linear regression model with a lag and two breakpoints. As there may be a lag between trends in suitability and trends in notified cases, we include a temporal term in the model to allow for a shift in the respective curves. Thus for each region, different sets of the constant and linear terms are fitted to different time periods. More formally,

$$\log(y_i + 1) = \alpha + \mathbb{I}(i \notin T)\alpha' + [b + \mathbb{I}(i \notin T)b']x_{i-l} \tag{4}$$

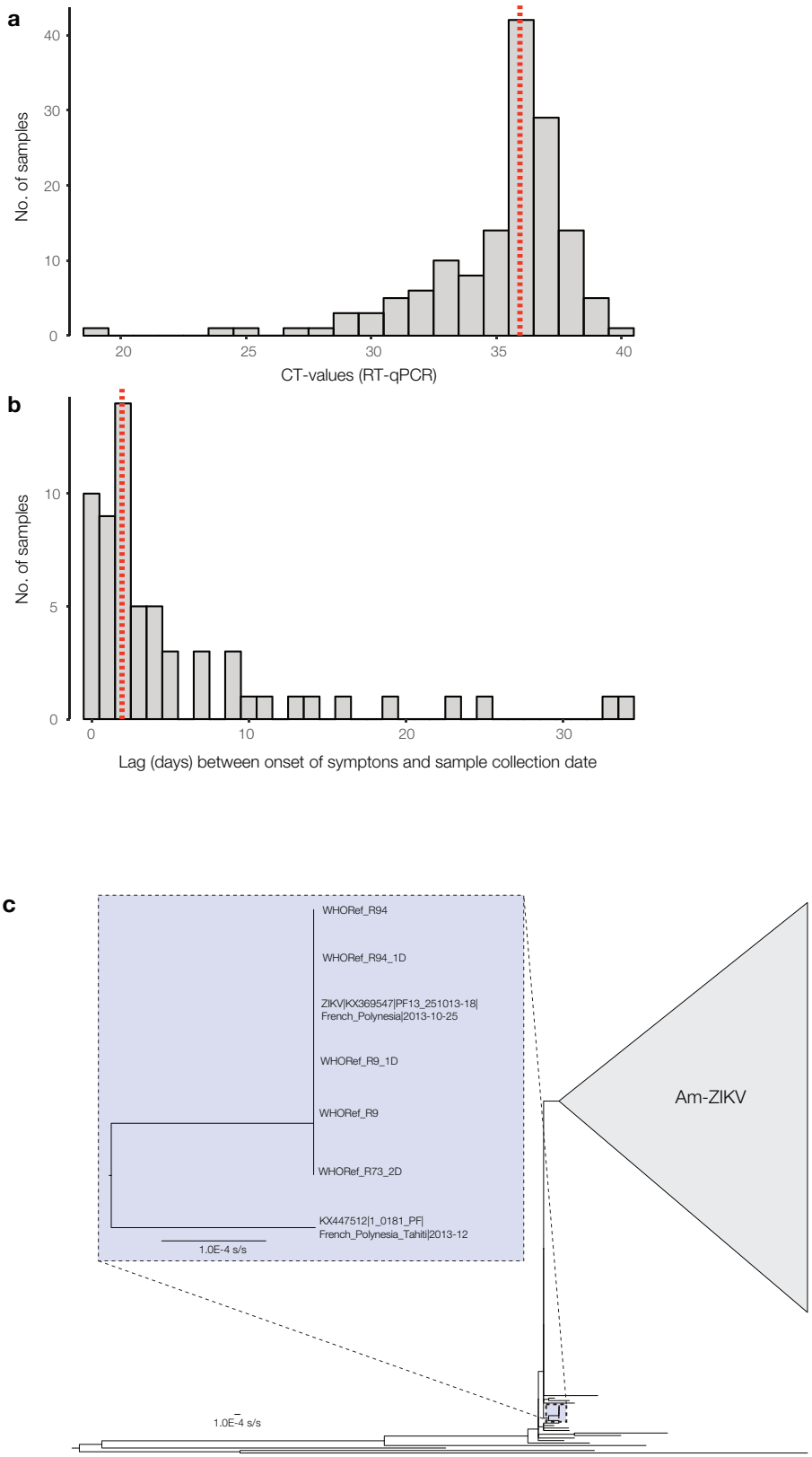
where y_i represents notified cases in a particular region in month i, x_i is the climatic suitability in that region in month i, l is the time lag that yields the highest correlation between y_i and x_i and T is the set of time indexes in the correlated region.

We then find the values of T and l that provide the highest adjusted R^2 by stepwise iterative optimization. For each value of T evaluated, the optimal value of l (that is, that which gives the highest adjusted-R^2 for the model above) is found by the optim function in R[58]. Climatic suitability values were only calculated for each month, so to calculate suitability values for any given point in time we interpolated between the monthly values using a linear function. We found no significant effect of residual autocorrelation in our data (Extended Data Fig. 7).

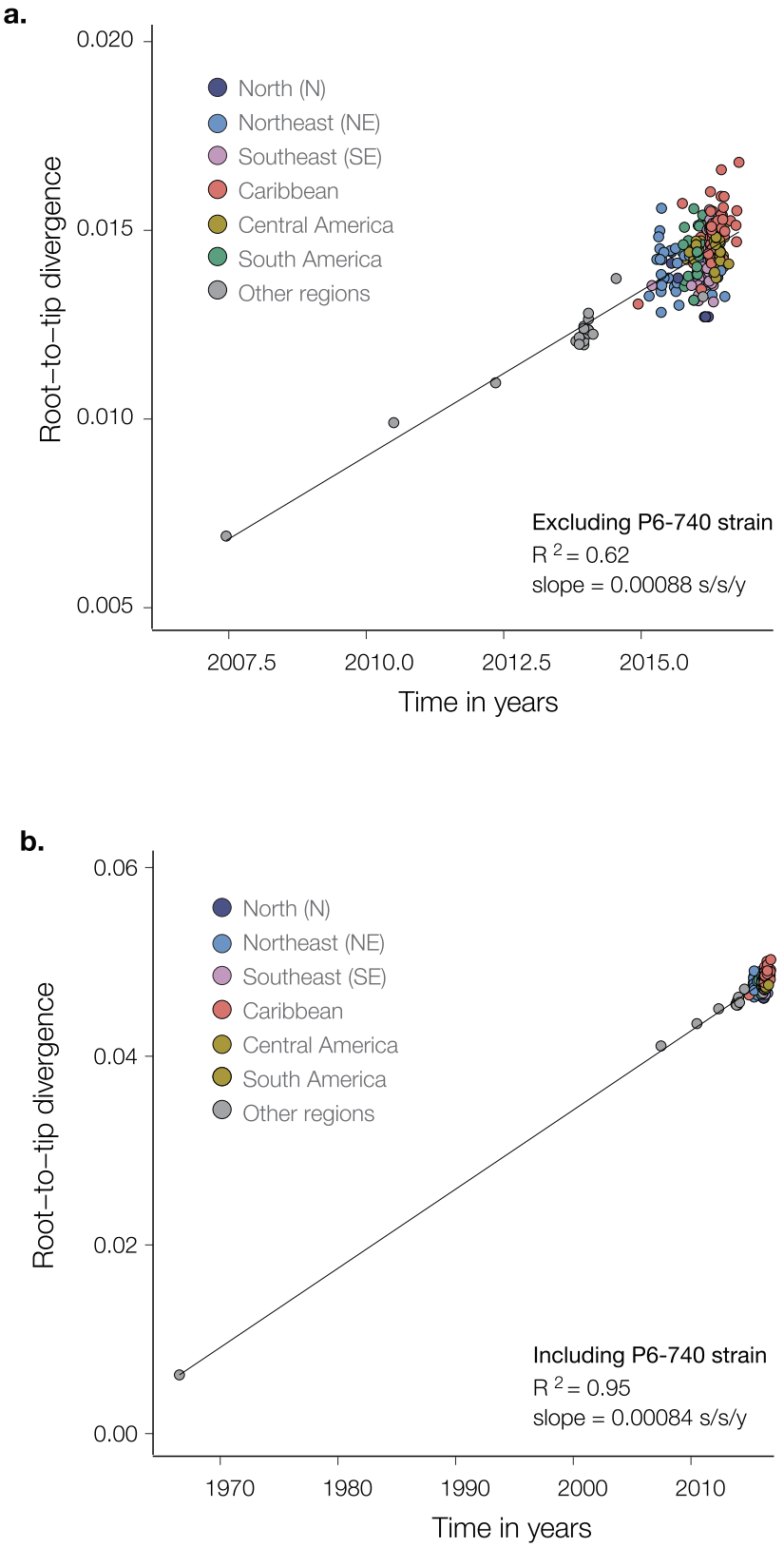
Data availability. Details of the primers and probes used here have been available at http://www.zibraproject.org since the beginning of the project. BEAST XML files, tree files, and sequence datasets analysed in this study are archived at https://github.com/zibraproject. New Brazilian sequences are available in GenBank under accession numbers KY558989–KY559032 and KY817930. New Colombian and Mexican sequences are available under accession numbers KY317936–KY317940 and KY606271– KY606274, respectively. See Extended Data Table 2 for further details.

35. 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).
36. Grubaugh, N. D. *et al.* Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* http://dx.doi.org/10.1038/nature22400 (2017).
37. Kozlov, A. M., Aberer, A. J. & Stamatakis, A. ExaML version 3: a tool for phylogenomic analyses on supercomputers. *Bioinformatics* **31,** 2577–2579 (2015).
38. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59,** 307–321 (2010).

39. 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).
40. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9,** 772 (2012).
41. Schierup, M. H. & Hein, J. Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156,** 879–891 (2000).
42. Faye, O. *et al.* Molecular evolution of Zika virus during its emergence in the 20th century. *PLoS Negl. Trop. Dis.* **8,** e2636 (2014).
43. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol* **1,** vev003 (2015).
44. Bruen, T. C., Philippe, H. & Bryant, D. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172,** 2665–2681 (2006).
45. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23,** 254–267 (2006).
46. Rambaut, A., Lam, T. T., Fagundes de Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2,** vew007 (2016).
47. 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).
48. Baele, G., Li, W. L., Drummond, A. J., Suchard, M. A. & Lemey, P. Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. *Mol. Biol. Evol.* **30,** 239–243 (2013).
49. Ferreira, M. A. R. & Suchard, M. A. Bayesian analysis of elapsed times in continuous-time Markov chains. *Can. J. Stat.* **36,** 355–368 (2008).
50. 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).
51. Kosakovsky Pond, S. L. & Frost, S. D. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **22,** 1208–1222 (2005).
52. Pond, S. L., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21,** 676–679 (2005).
53. Edwards, C. J. *et al.* Ancient hybridization and an Irish origin for the modern polar bear matriline. *Curr. Biol.* **21,** 1251–1258 (2011).
54. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10,** e1003537 (2014).
55. Minin, V. N. & Suchard, M. A. Fast, accurate and simulation-free stochastic mapping. *Phil. Trans. R. Soc. Lond. B* **363,** 3985–3995 (2008).
56. O'Brien, J. D., Minin, V. N. & Suchard, M. A. Learning to count: robust estimates for labeled distances between molecular sequences. *Mol. Biol. Evol.* **26,** 801–814 (2009).
57. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).
58. *R: A Language and Environment for Computing* (R Foundation for Statistical Computing, 2014).
59. Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **178,** 1505–1512 (2013).
60. Ferguson, N. M. *et al.* Countering the Zika epidemic in Latin America. *Science* **353,** 353–354 (2016).
61. Kraemer, M. U. *et al.* The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *eLife* **4,** e08347 (2015).
62. PAHO/WHO *Zika Epidemiological Update - Colombia (21 December 2016)* (World Health Organization, 2016).
63. PAHO/WHO *Zika Epidemiological Update - Mexico (20 December 2016)* (World Health Organization, 2016).
64. PAHO/WHO *Zika Epidemiological Update - Puerto Rico (20 December 2016)* (World Health Organization, 2016).

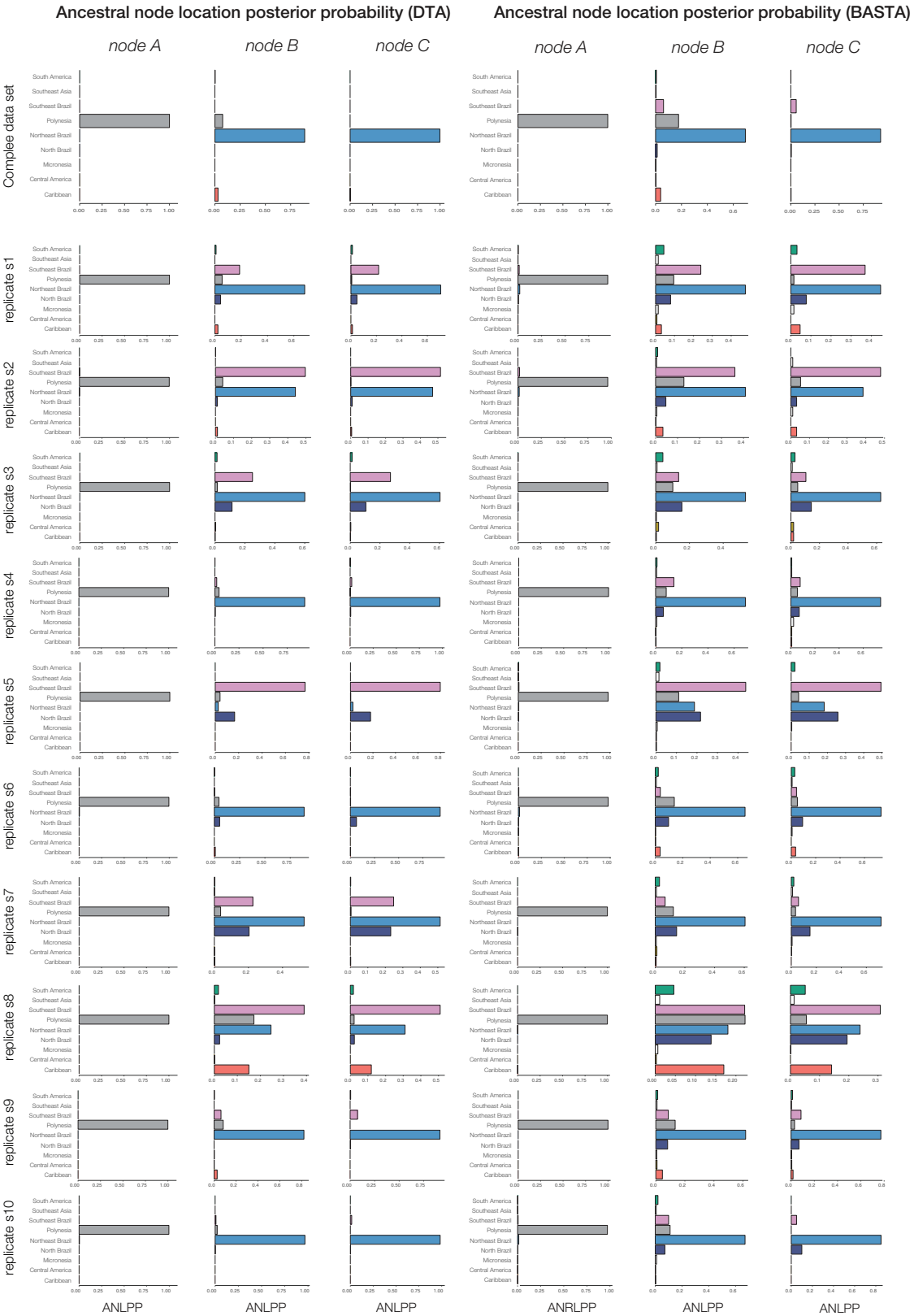


Extended Data Figure 1 | Characteristics of RT–qPCR$^+$ samples and validation of the ZIKV sequencing approach. a, Distribution of C_t values for the RT–qPCR$^+$ samples tested during the ZiBRA journey in Brazil ($n = 181$ samples; median $C_t = 35.96$). **b**, Distribution of the temporal lag between the date of onset of clinical symptoms and the date of sample collection of RT–qPCR$^+$ samples (median lag, 2 days). Red dashed lines represent the median of the distributions. **c**, Validation of sequencing approaches. A phylogeny of the ZIKV Asian genotype estimated using PhyML[38] is shown. The expanded clade highlighted in blue contains the WHO reference ZIKV sequence[20] (accession KX369547), which was generated using Illumina MiSeq. Sequences generated using MinION chemistries R9.4 2D, R9.4 1D, R9 1D, R9 2D and R7.3 2D contain no nucleotide differences and hence were also placed in this clade. Scale bars represent expected nucleotide substitutions per site (s/s).

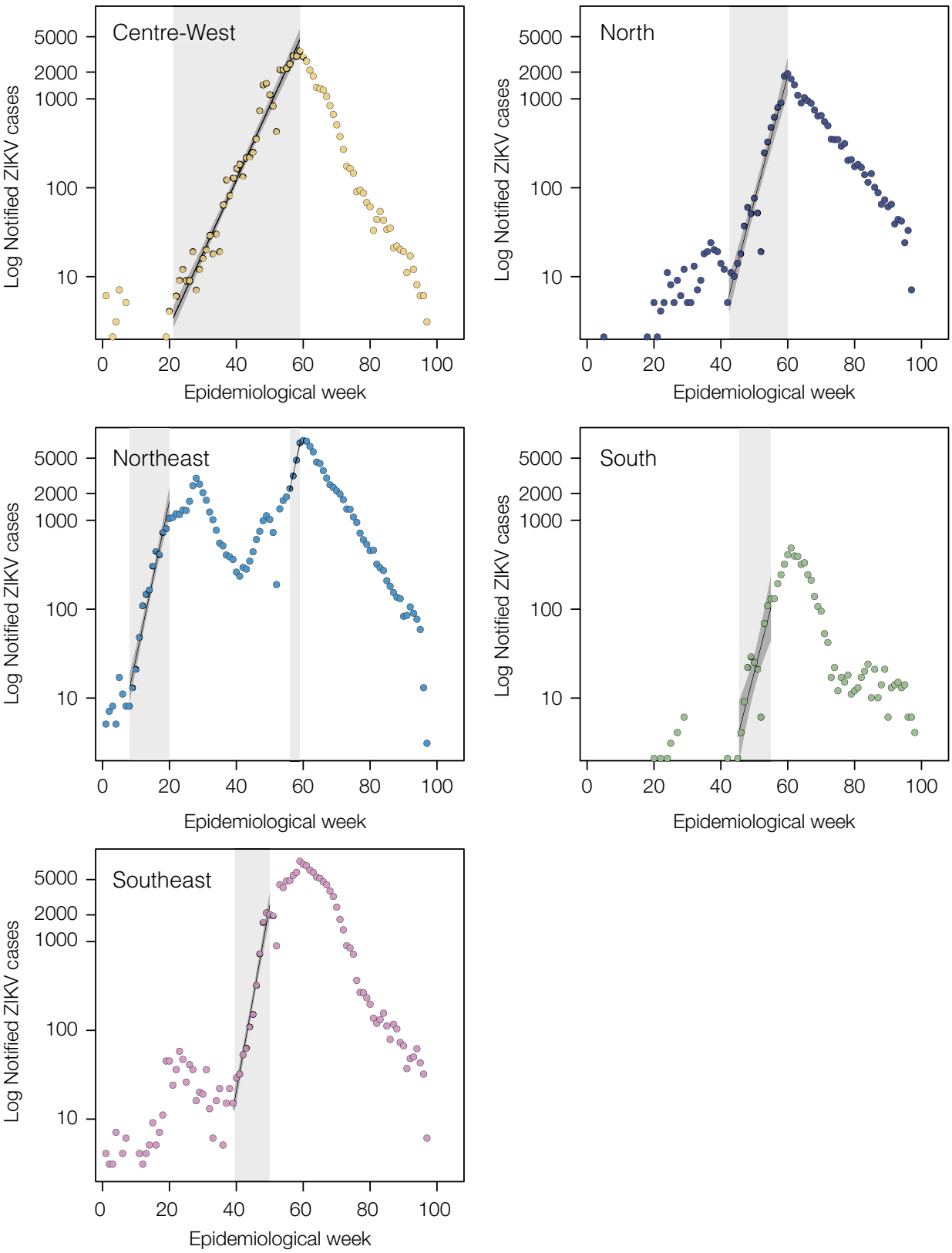


Extended Data Figure 2 | Temporal signal of the ZIKV Asian genotype. The correlation between sampling dates and genetic distances from the tips to the root of an ML tree, estimated using PhyML[38], was explored using TempEst[46]. **a**, Estimates for the dataset used in the phylogenetic analysis presented in Fig. 3c. **b**, Estimates for the same dataset with the addition of the P6-740 strain sampled in 1966 (accession number HQ234499).

Extended Data Figure 3 | A non-clock maximum likelihood phylogeny of our ZIKV dataset. Bootstrap branch support values are shown at each node. The phylogeny was estimated using PhyML[38]. Sequences generated in this study are highlighted in red. Scale bar represents expected nucleotide substitutions per site.

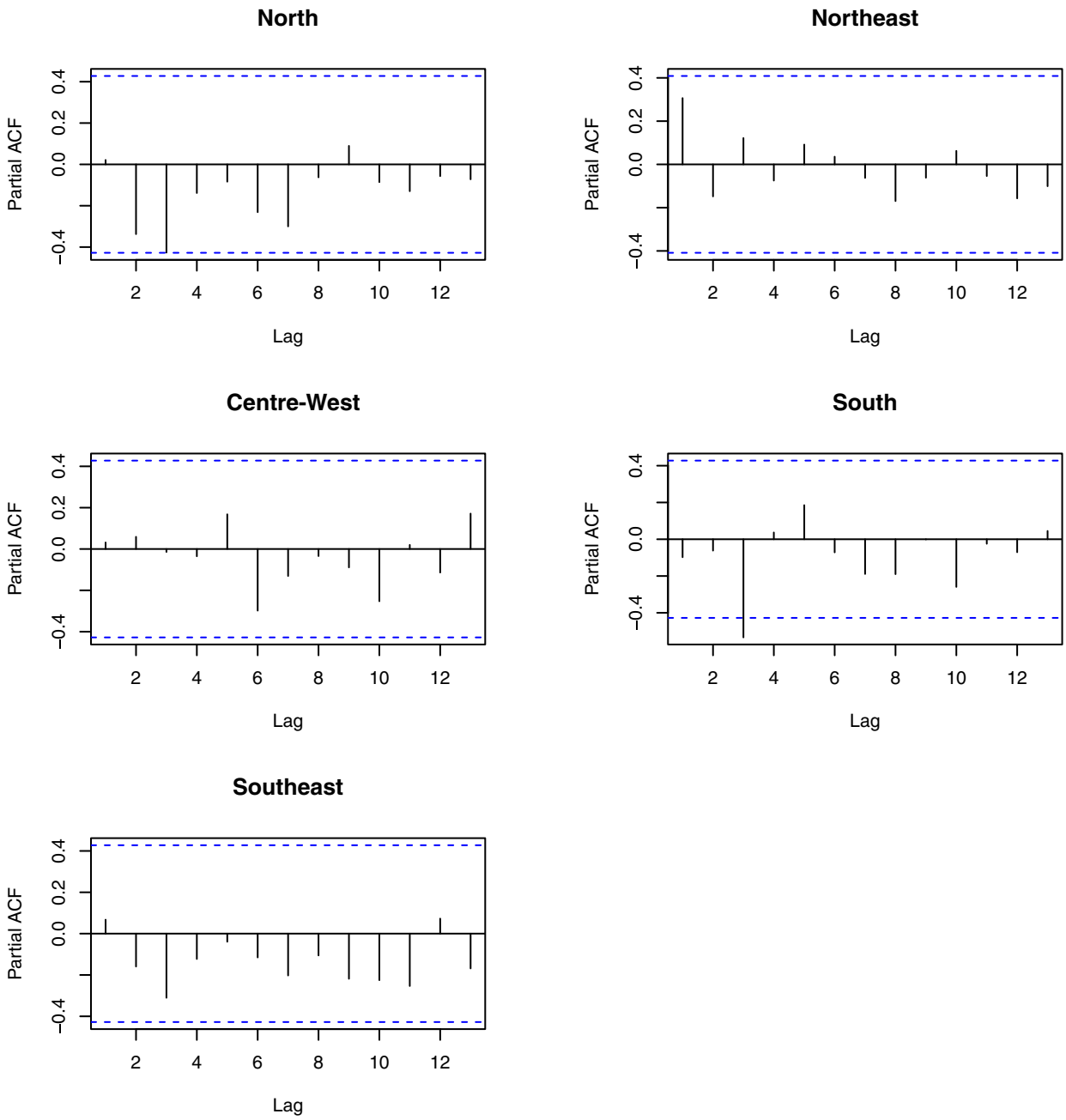


Extended Data Figure 4 | Ancestral node location posterior probabilities (ANLPP), for nodes A, B and C, estimated using the complete dataset (top row) and ten replicate subsampled datasets (other rows). See Methods for details. ANLPPs were calculated using two approaches: DTA = discrete trait analysis method[31] (left side columns) and BASTA = Bayesian structured coalescent approximation method[29] (right side columns). For each method, we employed an asymmetric model of location exchange to estimate ancestral node locations and to infer patterns of virus spread among regions.



Extended Data Figure 5 | Epidemic growth rates estimated from weekly ZIKV notified cases in Brazil. Time series show the number of ZIKV notified cases in each region of Brazil. Periods from which exponential growth were estimated are highlighted in grey.

Extended Data Figure 6 | Seasonal suitability for ZIKV transmission in the Americas. These maps were estimated by collating data on *Aedes* mosquitoes, temperature, relative humidity and precipitation, and are the basis of the trends in suitability for different regions shown in main text Figs 1 and 4. For method details, see refs 13, 6.



Extended Data Figure 7 | Partial autocorrelation functions for the linear model associating climatic suitability and ZIKV notified cases in each geographic region in Brazil. The residuals for the north, northeast, centre-west and southeast regions show no autocorrelation, whereas a small amount of autocorrelation cannot be excluded for the south region.

Extended Data Table 1 | Summary of the clinical samples tested in northeast Brazil and details of the epidemiological parameters obtained for each region in Brazil

(a)

Laboratory, Federal state	No. Positives / Tested (%)	Ct value (mean, min-max)	Collection lag (median, min-max)
LACEN, RN	27/335 (8.1%)	35.9 (18.6-39.1)	5 (4-16)
LACEN, PB	26/276 (9.4%)	35.7 (30.7-37.0)	6 (0-88)
FioCruz, PE	95/315 (30%)	34.6 (24.1-38.3)	2.5 (0-33)
LACEN, AL	16/140 (11%)	34.1 (27.1-40.2)	2 (0-3)
FioCruz, BA	17/264 (6.4%)	35.8 (24.7-39.2)	4 (0-228)

(b)

	N	NE	CW	S	SE
Correlated time period	12/2015 to 10/2016	7/2015 to 10/2016	9/2015 to 8/2016	6/2015 to 05/2016	11/2015 to 9/2016
***P*-value**	<0.0001	0.00013	<0.0001	<0.0001	<0.0001
Adjusted-R^2	0.929	0.8448	0.987	0.9543	0.953
Time lag (months)	1.27	0	1.12	1.19	1.33

(c)

Region	*R* (mean, CI), *g* =20 days	*R* (mean, CI), *g* =15 days	*R* (mean, CI), *g*=10 days	Growth rate (*r*, CI)
CW	1.71 (1.65-1.78)	1.46 (1.20-1.77)	1.29 (1.13-1.46)	0.027 (0.02-0.03)
N	2.48 (2.19-2.81)	1.98 (1.80-2.18)	1.58 (1.48-1.69)	0.046 (0.04-0.05)
NE, 1st	3.12 (2.69-3.60)	2.36 (2.11-2.63)	1.78 (1.65-1.91)	0.06 (0.05-0.07)
NE, 2nd	3.03 (2.74-3.36)	2.31 (2.14-2.49)	1.75 (1.66-1.84)	0.06 (0.05-0.06)
SE	3.85 (3.35-4.42)	2.77 (2.49-3.07)	1.98 (1.84-2.12)	0.07 (0.06-0.076)
S	2.57 (1.72-3.82)	2.04 (1.50-2.75)	1.61 (1.31-1.97)	0.05 (0.04-0.07)

a, Summary of the clinical samples tested ($n = 1,330$, of which 181 were RT–qPCR$^+$) by the ZiBRA mobile laboratory in June 2016, northeast Brazil. 84% of samples with known collection dates ($n = 698$ of 826) were from 2016. ZIKV notified cases were confirmed using RT–qPCR (see Methods). Collection lag represents the median time interval (in days) between the date of onset of clinical symptoms and date of sample collection (both dates available for $n = 219$) for all samples (including those that subsequently tested negative by RT–qPCR). Sample numbers in the FioCruz, PE row include RT–PCR$^+$ cases from Pernambuco generated at Fiocruz Pernambuco. **b**, Parameters of the model measuring the link between climatic vector suitability and notified ZIKV cases in different Brazilian regions. For each region, the table provides the estimated correlated time period (T), P value of the linear term of suitability in T, adjusted R^2 of the model, and time lag (l). **c**, For each region, estimates of the basic reproductive number (R) of ZIKV are shown for several values of generation time (g), together with the corresponding estimates of exponential growth rate (r) (per day) obtained from notified ZIKV case counts (see Extended Data Fig. 5). 1st: epidemic wave in 2015; 2nd: epidemic wave in 2016. CW, centre–west; N, north; NE, northeast; S, south; SE, southeast.

Extended Data Table 2 | Sequencing statistics

Accession Number	Sample ID	Aligned Reads	Consensus nucleotide bases (% of reference)	RT-qPCR Ct	Collection Date	Municipality	State
KY558989	ZBRA105	58128	9846 (92)	29.5	2015-02-23	João Câmara	RN
KY558990	ZBRC14	19111	8612 (81)	32.81	2016-01-15	Recife	PE
KY558991	ZBRC16	9161	7178 (67)	34.94	2016-01-19	Garanhuns	PE
KY558992	ZBRC18	7183	7459 (70)	35.14	2016-01-06	Caetes	PE
KY558993	ZBRC25	20533	5688 (53)	35.89	2016-01-18	Sanharo	PE
KY558994	ZBRC28	7905	8987 (84)	36.02	2016-01-18	Limoeiro	PE
KY558995	ZBRC301	20826	9843 (92)	31.99	2015-05-13	Paulista	PE
KY558996	ZBRC302	26331	10007 (94)	30.78	2015-05-13	Paulista	PE
KY558997	ZBRC303	12575	5873 (55)	32.81	2015-05-14	Olinda	PE
KY558998	ZBRC313	16530	9478 (89)	30.77	2015-06-15	Paulista	PE
KY558999	ZBRC319	17316	10565 (99)	24.07	2016-07-10	Olinda	PE
KY559000	ZBRC321	11434	8647 (81)	30.62	2015-08-09	Paulista	PE
KY559001	ZBRD103	13192	8380 (78)	29.09	2015-08-20	Murici	AL
KY559002	ZBRD107	77118	7415 (69)	30.31	2015-09-09	Maceió	AL
KY559003	ZBRD116	21211	9785 (92)	27.13	2015-08-28	Arapiraca	AL
KY559004	ZBRE69	2313	6866 (64)	24.72	2016-04-16	Feira de Santana	BA
KY559005	ZBRX1	21267	10559 (99)	25	2016-04-18	Ribeirão Preto	SP
KY559006	ZBRX2	24105	9961 (93)	32	2016-04-18	Ribeirão Preto	SP
KY559007	ZBRX4	14722	10563 (99)	26	2016-04-18	Ribeirão Preto	SP
KY559008	ZBRX6	12516	6893 (64)	33	2016-04-19	Ribeirão Preto	SP
KY559009	ZBRX7	10981	8563 (80)	33	2016-04-19	Ribeirão Preto	SP
KY559010	ZBRX8	7445	8702 (81)	33	2016-04-19	Ribeirão Preto	SP
KY559011	ZBRX11	21214	9379 (88)	31	2016-04-19	Ribeirão Preto	SP
KY559012	ZBRX12	19838	10305 (97)	31	2016-04-19	Ribeirão Preto	SP
KY559013	ZBRX13	11809	10564 (99)	21	2016-04-24	Ribeirão Preto	SP
KY559014	ZBRX14	5873	7469 (70)	33	2016-04-24	Ribeirão Preto	SP
KY559015	ZBRX15	20190	10563 (99)	27	2016-04-24	Ribeirão Preto	SP
KY559016	ZBRX16	9698	9027 (85)	32	2016-04-25	Ribeirão Preto	SP
KY559017	ZBRX100	5976	9609 (90)	28.5	2016-05-19	Ribeirão Preto	SP
KY559018	ZBRX102	13990	9508 (89)	33.91	2016-02-25	Porto Nacional	TO
KY559019	ZBRX103	17635	9514 (89)	36.76	2016-05-24	Araguaina	TO
KY559020	ZBRX106	29877	8458 (79)	32.36	2016-03-07	Palmas	TO
KY559021	ZRBX127	18914	10066 (94)	29.6	2016-03-10	Palmas	TO
KY559022	ZRBX128	18480	8650 (81)	28.79	2016-03-13	Palmas	TO
KY559023	ZBRX130	16667	9914 (93)	29.06	2016-03-22	Palmas	TO
KY559024	ZBRX137	15895	9767 (91)	34.83	2016-03-03	Palmas	TO
KY559025	ZBRY1	41036	8941 (84)	33.53	2016-01	Rio de Janeiro	RJ
KY559026	ZBRY4	27865	8433 (79)	34.21	2016-01	Rio de Janeiro	RJ
KY559027	ZBRY6	11779	10300 (97)	22.66	2016-01	Rio de Janeiro	RJ
KY559028	ZBRY12	4980	3061 (28)	33.66	2016-01	Rio de Janeiro	RJ
KY559029	ZBRY11	18530	5873 (55)	31.11	2016-01	Rio de Janeiro	RJ
KY559030	ZBRY10	14067	5712 (53)	30.84	2016-01	Rio de Janeiro	RJ
KY559031	ZBRY8	5708	9184 (86)	30.96	2016-01	Rio de Janeiro	RJ
KY559032	ZBRY7	7749	9018 (84)	28.07	2016-01	Rio de Janeiro	RJ
KY817930	ZBRY14	8040	5389 (50)	34.2	2016-02-15	Rio de Janeiro	RJ

Accession numbers, sample IDs, sequencing coverage, RT–qPCR values and epidemiological information for the samples from Brazil generated in this study. For the sequences from Rio de Janeiro state, alignments were performed against version 2 (KJ776791.2) of the genome reference; all other sequences used version 1 (KJ776791.1).

Extended Data Table 3 | Evolutionary analysis parameters and model selection results

(a)

Gene	Mean	Lower BCI	Upper BCI
C	0.96	0.72	1.20
prM	1.17	0.96	1.40
E	1.10	0.93	1.28
NS1	1.10	0.92	1.29
NS2A	1.26	1.04	1.52
NS2B	1.05	0.84	1.26
NS3	0.98	0.82	1.15
NS4A2K	1.26	1.02	1.55
NS4B	1.16	0.96	1.38
NS5	1.11	0.94	1.29

(b)

Clock	Coalescent	PS	SS
UCLN	Skyline	-32090.664	-32116.195
SC	Skyline	-32117.581	-32148.760
UCLN	Exponential	-32193.426	-32218.348
UCLN	Constant	-32206.219	-32234.196
SC	Constant	-32229.262	-32257.900
SC	Exponential	-32244.500	-32270.815

(c)

Clock model	Coalescent prior	*Node A* TMRCA (95% BCIs)	*Node B* TMRCA (95% BCIs)	*Node C* TMRCA (95% BCIs)
SC	Constant	2013.59 (2013.4,2013.77)	2013.83 (2013.6,2014.05)	2013.90 (2013.65,2014.12)
SC	Exponential	2013.59 (2013.38,2013.77)	2013.82 (2013.58,2014.04)	2013.89 (2013.65,2014.11)
SC	Skyline	2013.66 (2013.48,2013.81)	2013.93 (2013.74,2014.14)	2013.99 (2013.75,2014.18)
UCLN	Constant	2013.65 (2013.42,2013.84)	2013.91 (2013.63,2014.2)	2014.04 (2013.73,2014.32)
UCLN	Exponential	2013.66 (2013.45,2013.84)	2013.88 (2013.64,2014.13)	2014 (2013.73,2014.25)
UCLN	Skyline	2013.71 (2013.54,2013.85)	2014.03 (2013.76,2014.26)	2014.16 (2013.89,2014.41)

a, Estimated per-gene rates of evolution (mean and 95% Bayesian credible intervals (BCIs)) are shown in units of 10^{-3} substitutions per site per year. **b**, log-marginal likelihood estimates using the path-sampling (PS) and stepping-stone (SS) model selection approaches[47]. The overall ranking of the models is shown in parentheses for each estimator and the best-fitting combination is underscored. Two molecular clock models were tested here. SC, Strict clock model; UCLN, uncorrelated relaxed clock with log-normal distribution[46]. **c**, Estimated dates of nodes A, B and C (Fig. 3) under various different molecular clock and coalescent model combinations. BCI, Bayesian credible interval; SC, strict molecular clock model; TMRCA, time of the most recent common ancestor; UCLN, uncorrelated clock with log-normal distribution.