Infection, Genetics and Evolution 73 (2019) 33–43

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Time-scaled phylogeography of complete *Zika virus* genomes using discrete and continuous space diffusion models

Erika Ebranati[a,b,1], Carla Veo[a,b,1], Valentina Carta[a], Elena Percivalle[c], Francesca Rovida[c], Elena Rosanna Frati[b,d], Antonella Amendola[b,d], Massimo Ciccozzi[e], Elisabetta Tanzi[b,d], Massimo Galli[a,b], Fausto Baldanti[c], Gianguglielmo Zehender[a,b,*]

[a] Department of Biomedical and Clinical Sciences "L. Sacco", University of Milan, Milano, Italy
[b] CRC-Coordinated Research Center "EpiSoMI", University of Milan, Milano, Italy
[c] Molecular Virology Unit, Microbiology and Virology Department, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy
[d] Department of Biomedical Sciences for Health, University of Milan, Milano, Italy
[e] Unit of Medical Statistics and Molecular Epidemiology, University Campus Bio-Medico of Rome, Italy

ARTICLE INFO

Keywords:
Zika virus
Continuous phylogeography
Sourveillance
Phylodynamics

ABSTRACT

Zika virus (ZIKV), a vector-borne infectious agent that has recently been associated with neurological diseases and congenital microcephaly, was first reported in the Western hemisphere in early 2015.

A number of authors have reconstructed its epidemiological history using advanced phylogenetic approaches, and the majority of Zika phylogeography studies have used discrete diffusion models. Continuous space diffusion models make it possible to infer the possible origin of the virus in real space by reconstructing its ancestral location on the basis of geographical coordinates deduced from the latitude and longitude of the sampling locations. We analysed all the ZIKV complete genome isolates whose sampling times and localities were available in public databases at the time the study began, using a Bayesian approach for discrete and continuous phylogeographic reconstruction.

The discrete phylogeographic analysis suggested that ZIKV emerged to become endemic/epidemic in the first decade of the 1900s in the Ugandan rainforests, and then reached Western Africa and Asia between the 1930s and 1950s. After a long period of about 40 years, it spread to the Pacific islands and reached Brazil from French Polynesia. Continuous phylogeography of the American epidemic showed that the virus entered in north-eastern Brazil in late 2012 and started to spread in early 2013 from two high probability regions: one corresponding to the entire north-east Brazil and the second surrounding the city of Rio de Janeiro, in a mainly northwesterly direction to Central America, the north-western countries of south America and the Caribbean islands. Our data suggest its cryptic circulation in both French Polynesia and Brazil, thus raising questions about the mechanisms underlying its undetected persistence in the absence of a known animal reservoir, and underline the importance of continuous diffusion models in making more reliable phylogeographic reconstructions of emerging viruses.

1. Introduction

Zika virus (ZIKV) is an emerging arbovirus belonging to the *Flaviviridae* family, genus *Flavivirus*, and is closely phylogenetically related to other important mosquito-borne flaviviruses such as *Japanese encephalitis*, *West Nile* and *Dengue* viruses (Gubler et al., 2017). It was first discovered in a rhesus macaque monkey with fever kept in captivity in the Zika forest on the Entebbe peninsula in Uganda in 1947 (Buechler et al., 2017).

The viral genome is represented by a single-stranded positive-sense RNA molecule of 10.7 kbp that encodes a single polyprotein encompassing three structural proteins (the capsid, the precursor membrane and the envelope proteins) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5), which play essential roles in virus replication, virulence and secretion (Mittal et al., 2017). Phylogenetic studies of whole ZIKV genomes have revealed the existence of two major evolutionary lineages: one African and the other Asian (Simonin et al., 2017), encompassing also isolates from Pacific Islands

* Corresponding author at: Department of Biomedical and Clinical Sciences "L. Sacco", University of Milan, Via G.B. Grassi 74, 20157 Milan, Italy.
E-mail address: gianguglielmo.zehender@unimi.it (G. Zehender).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.meegid.2019.04.006
Received 5 December 2018; Received in revised form 1 April 2019; Accepted 6 April 2019
Available online 09 April 2019

and America.

ZIKV is naturally maintained by two distinct transmission cycles: a sylvatic cycle involving non-human primates and arboreal mosquitoes of the genus *Aedes*, and an urban cycle involving humans and urban mosquitoes (mainly *A. aegypti*). It can also be transmitted without vectors: vertically from an infected mother to her child during pregnancy, sexually (Barzon et al., 2016; D'Ortenzio et al., 2016), or by means of blood transfusions or exposure in a laboratory or healthcare setting (Lazear and Diamond, 2016). As with other arboviruses, about 80% of ZIKV-infected subjects are asymptomatic; symptomatic subjects most frequently experience flu-like syndrome, an itchy maculopapular rash and arthritis or arthralgia, but some cases of retro-orbital pain, headache, myalgia and vomiting have also been observed (Giovanetti et al., 2016). ZIKV can cause severe neurological complications such as Guillain-Barré syndrome and microcephaly in infants born to ZIKV-infected women, as demonstrated by the presence of the virus in the brain, placenta or serum of aborted fetuses and newborns with microcephaly (Sebastian et al., 2017). Since the early 1950s, ZIKV infection outbreaks have been reported in tropical Africa, south-eastern Asia and the Pacific islands. The virus was first isolated in Malaysia in 1969 and, later, in Indonesia (Marchette et al., 1969). In 2007, it caused the first large and well-characterised outbreak on Yap Island, a part of the Federated States of Micronesia (Duffy et al., 2009), and this was followed by a major epidemic in French Polynesia in 2013–2014 that affected > 28,000 people (11% of the population) (Musso et al., 2018) after which it spread to other neighbouring islands in the south Pacific.

In early 2015, the autochthonous transmission of ZIKV in the northeastern part of Brazil was the first reported description of the infection in the Americas (Calvet et al., 2016) and, by the end of the same year, ZIKV activity had expanded into at least 14 Brazilian states (https://www.paho.org). Other indigenous cases of ZIKV infection were detected in Colombia, Suriname, Paraguay and Venezuela in south America; Guatemala, El Salvador and Mexico in Central America, and Martinique and Puerto Rico in the Caribbean (Garcia-Luna et al., 2018). In early 2016, local outbreaks were confirmed in Guyana, Ecuador, Bolivia, Peru, Nicaragua, Curacao, Jamaica, Haiti, Santo Domingo and other Caribbean islands. In the first half of the same year, the virus was also detected in Argentina and Cuba and, finally, in the spring 2016, it reached the United States (Florida) (WHO data available at http://www.who.int/emergencies/zika-virus/history/en/). The epidemics started to decline in various American countries in the second half of 2016 and, although small local outbreaks were still being reported in 2017, their incidence was greatly reduced. Ultimately, a total of 48 countries in the Americas had > 540,000 autochthonous cases (> 200,000 in Brazil) and there were about 2610 reported congenital infections (2300 in Brazil) [WHO-PAHO: Regional Zika Epidemiological Update (Americas) August 25, 2017, available at https://www.paho.org].

ZIKV has now become endemic not only in Souh America and Caribbean, but also in several Pacific islands (American Samoa, the Federated States of Micronesia, Fiji, Marshall Islands, New Caledonia, Samoa and Tonga) (Calvez et al., 2018). In addition, there have been an increasing number of travel-related cases in non-endemic countries such as Australia, Belgium, Canada, China, France, Portugal, Spain, Switzerland and The Netherlands (De Smet et al., 2016).

Several authors have attempted to study the dynamics of Zika virus infection through discrete phylogeographical analysis (Boskova et al., 2018; Faria et al., 2017; Giovanetti et al., 2016; Liang et al., 2017; Metsky et al., 2017; Pettersson et al., 2018).

In addition to classical discrete phylogeographical methods, new continuous diffusion models based on Brownian or random walk diffusion models have been developed that can infer ancestral states on the basis of the coordinates of a two-dimensional space identifying the tips of the tree (sampling location). These models allow a more realistic reconstruction of spatial movements because, unlike discrete models, they do not necessarily need the ancestral location to be represented in the sampling location set (Lemey et al., 2010). The differences between discrete and continuous phylogeographic models have been efficiently described in previous reviews (Bloomquist et al., 2010; Faria et al., 2017).

The aim of this study was to infer the origin and dispersion routes of ZIKV in the world using a classical discrete method and to reconstruct the recent epidemic in the Americas using a continuous phylogeographical method to better describe the local spread of ZIKV and to make hypothesis about the eco/epidemiology of the virus.

2. Materials and methods

2.1. Patients and datasets

The study was conducted using 135 complete viral genome sequences retrieved from public databases. These sequences were derived from mosquitoes ($n = 21$), a sentinel rhesus monkey ($n = 6$), and human samples ($n = 108$). The sequences were isolated in various countries of the world and retrieved from GenBank (at http://www.ncbi.nlm.nih.gov./genbank/); only the sequences with a known location and sampling date were considered.

The sampling period ranged from 1947 to 2016, and the sampling locations ranged from the Central African Republic (*CF*, $n = 4$) to Nigeria (NG, $n = 2$), Senegal (SN, $n = 10$), Uganda (UG, $n = 6$), southeast Asia (SEA, including Malaysia n = 4, Cambodia n = 2, Thailand $n = 3$, and Singapore n = 2), French Polynesia (FP, $n = 11$), Mexico (MX, $n = 8$), Honduras (HN, n = 6), Guatemala (GT, n = 3), Panama (PA, $n = 4$), Venezuela (VE, n = 8), Colombia (CO, n = 4), Ecuador (EC, $n = 2$), Brazil (BR, $n = 28$), Suriname (SR, n = 2), the Western Pacific (WP, including American Samoa n = 6, Tonga $n = 1$, the Philippines n = 1, Micronesia n = 1), the Antilles (ANT, including Puerto Rico n = 4, Haiti n = 2, Dominican Republic n = 8, Martinique n = 1, Cuba n = 2). Two of the 28 Brazilian samples were from tourists returning to Italy who became infected in Bahia State.

The sequences were selected on the basis of the following criteria: i) they had to have been published in peer-reviewed journals; ii) their non-recombinant subtype assignment had to be certain; and iii) the city/state of origin and year of sampling had to be known and clearly established in the original publication. The origins and characteristics of the Zika strains dataset are summarised in Supplementary Table 1.

2.2. Ethics statement

Informed consent was obtained according to Italian law (art.13 D.Lgs 196/2003) as well as approval of the Institutional review board of Fondazione IRCCS Policlinico San Matteo on the use of residual biological specimens (IRB Protocol 20100000348).

2.3. Whole genome characterisation by means of next-generation sequencing

The whole ZIKV genome sequence of one human isolate (an Italian semen sample from a traveller coming from Bahia State in Brazil) was previously obtained by Sanger methodology and deposited in GenBank database with the accession number KY003154. Subsequently, in our laboratory, we amplified the whole genome by using the sequence-indipendent single-primer amplification (SISPA method) and re sequenced it by NGS method (Djikeng et al., 2008).

ZIKV was isolated on Vero E6 cells; the RNA was prepared by extracting it from the cell culture supernatant and then it was reverse-transcribed using the random primer FR26RV-N (5'GCCGGAGCTCTGC AGATATCNNNNNN3') at a concentration of 10 µM. Viral cDNA was denatured at 94 °C for three minutes, and chilled on ice for two minutes. Five units of Klenow fragment (New England Biolabs, Ipswich, MA) were directly added to the reaction to perform the second strand cDNA

synthesis. The incubation was carried out at 37 °C for one hour, and at 75 °C for 10 min.

Next, 5 uL of double-stranded DNA were added to a PCR master mix containing 5 μL of 10× AccuPrime PCR buffer I, 0.2 μL of AccuPrime Taq DNA Polymerase high fidelity, 4 uL of 10 μM FR20RV (5'GCCGG AGCTCTGCAGATATC3') and 35.8 μL of water. The incubation was performed under the following thermal conditions: 94 °C for two minutes, 40 cycles of 94 °C for 30 s, 55 °C for one minute and 68 °C for three minutes.

The PCR product was purified and quantified using a TECAN plate reader. The sample was diluted to an initial concentration of 0.2 ng/μL in accordance with the Illumina protocol, and 1 ng was used for the library preparation (Nextera XT sample preparation Kit, Illumina Inc., San Diego, California, USA).

Genomic libraries were sequenced on the Illumina MiSeq platform (Illumina, Inc.) with 2 × 151 base pairs paired-end runs. Finally, we evaluated the obtained reads for sequence quality and read-pair length using FastQC ver. 0.11.5.

The reads were assembled using Geneious software v. 11.1.5 (Biomatters, New Zealand) and re-sequencing analysis was performed with the reference virus (KY003154).

2.4. Recombination detection

In order to identify recombinant strains and exclude them from the analysis, we used the RDP4 package, which allows the identification of potential recombinant sequences and their parents (major and minor). It uses seven different methods: RDP (Martin et al., 2015), BOOTSCAN (Martin et al., 2005), CHIMAERA (Posada and Crandall, 2001), SISCAN (Gibbs et al., 2000), GENCONV (Padidam et al., 1999), 3SEQ (Boni et al., 2007) and MAXCHI (Smith, 1992), each of which has a highest acceptable p value of 0.05 and Bonferroni's correction for multiple comparisons The sequences indicated as being recombinant by at least three of these methods were excluded from further analysis.

We also screened our alignment using Genetic Algorithm Recombination Detection (GARD) software in order to detect any sequences involved in putative recombinations, and define the number and location of breakpoints (Kosakovsky Pond et al., 2006).

2.5. Likelihood mapping

The phylogenetic signal of the complete genome dataset was investigated by means of the likelihood mapping (LM) analysis of 10,000 random quartets generated using TreePuzzle (Strimmer and von Haeseler, 1997). Groups of four randomly chosen sequences (quartets) were evaluated and, for each quartet, the three possible unrooted trees were reconstructed using the maximum likelihood approach under the selected substitution model that was the General Time Reversible(GTR) with gamma distributed rates among sites. The posterior probabilities of each tree were then plotted on a triangular surface so that fully resolved trees fell into the corners, and the unresolved quartets in the centre of the triangle (a star-tree). When using this strategy, if > 30% of the dots fall into the centre of the triangle, the data are considered unreliable for the purposes of phylogenetic inference.

2.6. Phylogenetic reconstruction

The sequences were aligned using ClustalX software (Jeanmougin et al., 1998) followed by manual editing using Bioedit software v. 7.2.6, and the best fitting nucleotide substitution model was tested by means of the hierarchical likelihood ratio test (LRT) implemented in J Modeltest software (Posada, 2008). The selected model was GTR with gamma distributed rates among sites.

The phylogeny of the complete genome was reconstructed using a maximum likelihood approach and the new hill-climbing algorithm implemented in PhyML v.3.0. The reliability of the observed clades was established on the basis of internal node bootstrap values of ≥70% (after 200 replicates) (Guindon et al., 2010).

The phylogeny of the complete genome was also reconstructed using a Bayesian Markov Chain Monte Carlo (MCMC) method (Beast v. 1.8.4 freely available at http://beast.bio.ed.ac.uk). The reliability of the observed clades was established on the basis of posterior probabilities values with significance levels of ≥0.7.

The evolutionary rates were estimated under strict and relaxed (with an uncorrelated log normal rate distribution) clock conditions.

As coalescent priors, three parametric demographic models of population growth (constant size, exponential growth and logistic growth) the Bayesian SkyGrid, the Bayesian skyline plot (BSP) and the GMRF Bayesian Skyride were compared (Drummond et al., 2005). The best fitting models were selected using the BF implemented in Beast. In accordance with Kass and Raftery, the strength of the evidence against H_0 was evaluated as 2lnBF < 2 = no evidence; 2–6 = weak evidence; 6–10 = strong evidence, and > 10 = very strong evidence. A negative 2lnBF indicates evidence in favour of H_0. Only values of ≥6 were considered significant (Kass and Raftery, 1995).

We also used path sampling (PS) and stepping stone sampling (SS) to improve the accuracy of model selection (Baele et al., 2012).

The chains were run for 250 million generations until reaching convergence and sampled every 25,000 steps. Convergence was assessed by estimating the effective sampling size (ESS ≥ 200) after a 10% burn-in, using Tracer software version 1.6 (http://tree.bio.ed.ac.uk/software/tracer/). All of the parameters had an ESS of > 200. Uncertainty in the estimates was indicated by 95% highest posterior density (95% HPD) intervals.

The TMRCA estimates were expressed as the median and 95% HPD years before the most recent sampling date, which corresponded to 2016 in this study.

2.7. Root-to-tip regression analysis

In order to verify the correlations between time and genetic distances and identify the correct root under the hypothesis of proportionality between them, we used Tempest software and the ML tree to make a root-to-tip regression analysis (Rambaut et al., 2016).

2.8. Bayesian phylogeographic analyses

2.8.1. Discrete phylogeographic analysis

An improvement to Beast allows an ancestral reconstruction of discrete states in the Bayesian framework described above in which the spatial diffusion of the time-scaled genealogy is modelled as a continuous-time Markov chain process over discrete sampling locations. A Bayesian stochastic search variable selection (BSSVS) approach, which allows the exchange rates in the CTMC to be zero with some prior probability, was used in order to find a minimal (parsimonious) set of rates explaining the diffusions in the phylogeny. Comparing the posterior to prior odds that individual rates are zero provides a Bayes factor test to identify the rates contributing to the migration pathway, which were calculated as described elsewhere. Rates yielding a BF of > 3 were considered significant (Lemey et al., 2009).

The obtained trees were summarised in a maximum clade credibility tree using the Tree Annotator program included in the Beast package, choosing the tree with the maximum product of posterior probabilities (maximum clade credibility: MCC) after a 10% burn-in. The most probable location of each node was highlighted by labelling the branches with different state colours. In order to visualize diffusion rates over time, it is possible to convert the location-annotated MCC tree to a keyhole mark-up language file (KML) suitable for viewing with georeferencing software.

In order to visualize diffusion rates over time, it is also possible to render the location-annotated MCC tree to a GeoJSON data format suitable for viewing with georeferencing software. The new SPREAD3

analysis tool was used, the MCC tree was converted to a JavaScript object notation (JSON) file and the visualization was rendered using a Data Driven Document (D3) library (Bielejec et al., 2016).

2.8.2. Continuous phylogeographic analysis

In order to study the spread of ZIKV in more detail, a continuous space phylogeographical analysis was made using American isolates.

American ZIKV epidemics were investigated in continuous space using Beast v. 1.8.4. The unknown coordinates were estimated under a strict Brownian diffusion model, and compared with two relaxed random walk (RRW) models relaxing the diffusion rate constancy assumption that respectively assumed the gamma and Cauchy distribution of diffusion rates over the phylogeny (Lemey et al., 2010). Bayes factor comparisons of the models were made by estimating marginal likelihood using path sampling (PS) and stepping stone approaches (Baele et al., 2012). The phylogeny was spatially projected and converted into KML in order to visualize dispersal over time. Uncertainties in the ancestral location estimates were represented by KML polygons delimiting the high-probability regions.

3. Results

3.1. Illumina paired-end sequencing

The raw data reads with quality value QV > 20 were filtered by excluding contaminants such as adapters, the ambiguous "N" nucleotides and low-quality sequences using trimming options implemented in Geneious software. After trimming the raw data, a total of 178,846 filtered clean reads were obtained with a 76,023× coverage.

3.2. Recombination analysis

Both the GARD and RDP programs confirmed the presence of recombination in the final alignment. In particular, RDP analysis (Supplementary Table 2) detected four genomes showing a total of 12 significant recombination events corresponding to isolates from Senegal (12SN@68, 13SN@97, 14SN@01 and 15SN@01). For this reason, these sequences were removed from the definitive data set.

3.3. Likelihood mapping analysis

The presence of phylogenetic noise was investigated using LM analysis. The complete genome data set gave satisfactory results as 7.8% of the dots fell into the central area of the triangles and 87.4% at the corners, thus suggesting that the alignment contained sufficient phylogenetic information (Supplementary Fig. 1).

3.4. Phylogenetic analysis

ML analysis of the whole genomes showed two statistically supported clades (bootstrap = 1000) corresponding to the previously described African (AF) and Asian (AS) clades (Fig. 1). Two different subclades could be distinguished within the African clade: the first (eastern central African, EC) sub-clade included the original 1947 Ugandan isolates, four genomes from the Republic of Central Africa obtained between 1968 and 1980, and two Senegalese isolates obtained in 2001 (each of them grouping together on a geographical basis); the second (Western African, W) sub-clade encompassed the majority of the Senegalese isolates obtained between 1968 and 1997 and two Nigerian strains obtained in 1968, all significantly segregating on the basis of their geographical origin.

The Asian/Pacific Ocean clade (AS) included two geographically distinct monophyletic groups: the first including all of the isolates from Malaysia 1966 (MY), and the second and largest group (AWP) including all of the other Asian, Western Pacific Ocean, Polynesian and American isolates connected by a long branch (indicating a bottleneck) to the Malaysian clade. The American sub-clade was statistically sustained (bootstrap = 1000). Analysis of the single genes highly supported these clades and sub-clades (Fig. 1).

3.5. Root-to-tip regression analysis

Analysis of the unrooted ML tree of the entire dataset without the recombinant sequences, showed a very strong association between genetic distances and sampling dates ($R^2 = 0.93$, correlation coefficient = 0.97), thus confirming the suitability of the dataset for molecular clock analysis. Tempest also located the best tree root within the branch connecting the African and Asian clades.

Interestingly, separate analysis of each gene (Table 1) showed similar results in all the datasets, with the weakest temporal signal being obtained using the Membrane dataset ($R^2 = 0.5$).

3.6. Evolutionary rates, tMRCA estimates and Bayesian phylogeography

The evolutionary rates, tMRCAs and phylogeography were co-estimated using a Bayesian framework implemented in Beast (v. 1.8.4). The comparison by Bayes factor of the marginal likelihoods obtained by applying a strict or relaxed molecular clock under five different coalescent models (2lnBF GMRF Bayesian Skyride *vs* BSP = − 811.92; 2lnBF constant *vs* BSP = − 68.31; 2lnBF exponential growth *vs* BSP = −1092.56; 2lnBF Bayesian Skygrid *vs* BSP = − 748.1); under a log-normal relaxed clock (2lnBF strict *vs* relaxed clock = −466.62) showed that the favoured models were the relaxed molecular clock with uncorrelated log-normal rate distribution and the Bayesian skyline plot, the less stringent demographic model. The same model has been confirmed by using path sampling model selection (PS) and stepping stone sampling (SS) model selection (Table 2). Under these conditions, we estimated a mean substitution rate for the entire viral genome of 8×10^{-4} (95% HPD 6.4–9.7 $\times 10^{-4}$) subs/site/year (Table 1 shows the mean estimates for each single gene).

The tree-root tMRCA (Fig. 2) was estimated to be an average of 114.2 (95% HPD 84–148) years before the present, corresponding to the year 1902 (95% HPD 1868–1932). The MRCA of the eastern Central African clade was placed in 1930 (95% HPD = 1918–1940), whereas that of the West African sub-clade originated later, in 1954 (95% HPD = 1944–1963). The first Asian node (corresponding to the Malaysian group) dated back to 1958 (95% HPD 1951–1965), and was connected by a long branch to the largest AWP sub-clade whose tMRCA dated back to 1998 (95% HPD 1993–2003). A further highly significant sub-clade (pp = 1) included all of the American strains, which had an estimated mean tMRCA of 3.1 years ago (95% HPD 2.8–3.5), corresponding to the year 2013 (2012–2014). In general, the isolates of this AWP clade also tended to segregate significantly on the basis of their geographical origin (Table 3).

Analysis of migration flows showed 19 (95% HPD = 17–21) non-zero rates between different localities, all of which were significant at BF analysis (BF > 3). The suggested dispersion pathway summarised in Fig. 3 showed that the currently circulating ZIKV strains shared a common ancestor that existed in eastern Central Africa in the first decades of the 1900s, and spread to Western Africa in the 1950s. In the same period, it spread to Asia (Malaysia) and the Asian strain reached the Pacific islands at least twice: in the first decade of the 2000s and in 2012, when it spread to French Polynesia. Finally, from French Polynesia, it reached Brazil in 2013 to start a flow that subsequently reached a several number of central and south American regions.

Considering the discrete phylogeographic tree and limiting the analysis to the Asian clade without the Malaysian strains, we observed a total of 14 highly significant (0.9 < pp. < 1) monophyletic groups including more than two isolates (a median of five isolates for clade, range 3–9). The clades were strongly defined on a geographical basis (Supplementary Table 3): five were Brazilian clades (three including only Brazilian isolates, and two including mixed isolates from Brazil

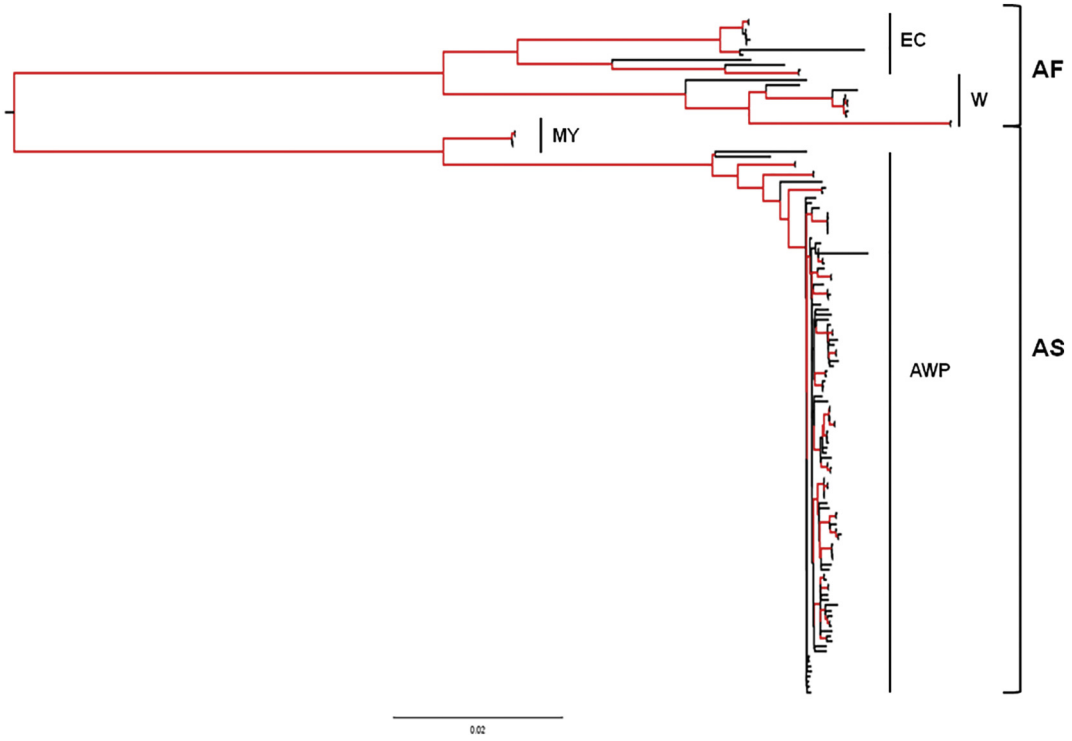

Fig. 1. Maximum likelihood tree of the 133 *Zika virus* complete genome sequences. The significant posterior probabilities (pp ≥ 0.7) of the corresponding nodes have been coloured in red and the main clades have been highlighted. The scale bar indicates 2% of nucleotide divergence. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Comparison between regression analysis and phylogeographic analysis for each single *Zika virus* gene.

	Root-to-tip					Bayesian analysis	
	Slope	tMRCA[a]	Correlation	R Squared	Residual mean	E.R[b] MEAN	tMRCA[c]
	($*10^{-3}$)		Coefficient		Squared	(95% HPD[c] lower-upper)	
Capsid	1.19	1878.3	0.97	0.94	$2.97*E^{-5}$	1.03 (0.42–1.6)	1936.5
Membrane	1.02	1885.4	0.73	0.54	$3.2*E^{-4}$	2.2 (1.06–3.5)	1944.8
Envelope	0.85	1802.2	0.91	0.84	$4.98*E^{-5}$	1.7 (0.95–2.4)	1940.6
NS1	0.45	1806.1	0.92	0.84	$1.32*E^{-5}$	1.4 (0.92–2)	1944.3
NS2	1.18	1811.3	0.95	0.9	$5.48*E^{-5}$	1 (0.46–1.6)	1924.9
NS3	0.67	1863.2	0.97	0.94	$9.99*E^{-6}$	0.85 (0.53–1.2)	1924.7
NS4	0.59	1826.9	0.95	0.9	$1.33*E^{-5}$	0.86 (0.58–1.3)	1910.1
NS5	0.63	1835.4	0.97	0.93	$1.03* E^{-5}$	0.93 (0.63–1.3)	1913.6

[a] tMRCA: time of the most Recent Common Ancestor.
[b] E.R: Evolutionary Rate.
[c] HPD: Highest posterior density, substitutions/site/year ($*10^{-3}$).

and Ecuador or Italy); two included Venezuelan isolates in one clade with isolates from Colombia and the other with one Dominican sequence; two were pure Central American clades (one from Mexico and the other from Panama) and two others were from the Caribbean (one pure from Puerto Rico and one mixed from Cuba and Santo Domingo, with one strain from Mexico). The earliest tMRCA was that of the French Polynesia clade, followed by the Brazilian clade, the Central American clades and, finally the clades from Antilles and Venezuela.

In order to reconstruct the spread of ZIKV in the new world in more detail and avoid the limitations caused by their arbitrary grouping into discrete localities, we used a continuous phylogeographic model based on the geographical coordinates of the sampling localities. This analysis only considered data subset of American isolates. Comparison of the strict Brownian diffusion model (assuming a homogeneous diffusion rate over the phylogeny) with two RRW models (assuming different diffusion rates on each branch of the tree) by the BF test showed that a log-normal RRW diffusion rate fitted the data better than the other models (Cauchy distribution RRW *vs* homogenous BD: 2lnBF = 639.34 by PS and 175.78 by SS; Cauchy distribution RRW *vs* log-normal RRW: 2lnBF = 1335.32 by PS and 10.64 by SS). On the basis of this continuous phylogeographical reconstruction, the tree root was placed between the coordinates −41.13 E of longitude and − 9.53 N of latitude, corresponding to a location in the state of Bahia in north-east Brazil, close its border with the two other states of Pernambuco and Piauì (see the animated visualization in Supplementary Video 1). Two high probability (80% HPD) regions (Fig. 4) were identified almost simultaneously at the beginning of the epidemic (2013): the first (A in Fig. 4, panel 1) was a large ellipse with a major axis of about 1700 km and a minor axis of about 700 km that included the tree-root and encompassed north-east Brazil (the nine states of Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte, and Sergipe); the second (B in Fig. 4, panel 1) was a smaller circle with a

Table 2
Model selection using path sampling and stepping stone sampling under strict and uncorrelated log normal clock.

Clock	Models	Complete genome PS[a]	Complete genome SS[b]
Strict	Constant	−37,209,53	−37,259,37
Strict	Exponential	−37,204,45	−37,271,32
Strict	Skygrid	−37,208,28	−37,265,18
Strict	Skyline	−37,173	−37,235,26
Strict	Skyride	−37,417,47	−37,499,97
UCLN[c]	Constant	−37,148,91	−37,218,2
UCLN[c]	Exponential	−37,153,79	−37,225,78
UCLN[c]	Skygrid	−37,151,84	−37,196
UCLN[c]	Skyline	−37,100,32	−37,163,92
UCLN[c]	Skyride	−37,362,07	−37,411,22

[a] Path sampling.
[b] Stepping stone sampling.
[c] Uncorrelated log normal.

diameter of about 400 km, surrounding the metropolitan area of Rio de Janeiro in south-east Brazil. These two areas were the origin of different migration flows, corresponding to the branches and clades highlighted in the continuous phylogeographic tree (Fig. 4, panels 1–4, and Fig. 5). Area A gave rise to two main migration pathways: the first (corresponding to clade A1 in the tree) initially spread across north-east Brazil and subsequently reached Santo Domingo (2015), Cuba and Ecuador (2016); the second (corresponding to clade A2, and not significantly segregated from the root of the tree) first reached the island of Haiti (2014) and then spread to Venezuela (2016). Area B was the origin of three main migratory flows: the first (B1) reached Suriname in 2013, and then proceeded towards Porto Rico (2015); the second (corresponding to clade B2) spread to an uncertainty region encompassing Panama and Colombia in 2013 and subsequently (2015/2016) dispersed east towards the Caribbean (Martinica, Santo Domingo) and Venezuela, and north towards Mexico; and the third (B3) reached Honduras and Guatemala before spreading throughout Central America and reaching Mexico in the last year (2016).

Table 4 shows the geographical coordinates of the localities and tMRCA estimates for each of the main clades (root, A1-2, B1-3). The estimated average tree-root tMRCA was 3.7 years ago, corresponding to October 2012 (95% HPD December 2011 to August 2013), and the estimated average tMRCAs of the main clades were between March (clade A2) and August 2013.

Within a few months, ZIKV had spread to the entire continent. It followed a mainly north-westerly pathway at the beginning of the epidemic but, between late 2014 and the beginning of 2015, went in various directions (along an east/west axis between central America and the Caribbean, and even in a south-easterly direction), thus indicating wider viral dispersion throughout the region. The estimated overall diffusion rate was as much as 760 km/year (between about 600 and 900 km/year).

4. Discussion

A number of authors have recently attempted to estimate the evolutionary dynamics of ZIKV in the Americas with the aim of reconstructing the most probable origin of the epidemic, the time of its entry into the Americas, and the diffusion pathways that led to its spread across the continent in such a short time. Some of these studies used partial coding sequences (Giovanetti et al., 2016; Liang et al., 2017), and others whole viral genomes (Faria et al., 2017; Metsky et al., 2017). Some authors included in their analyses all of the isolates

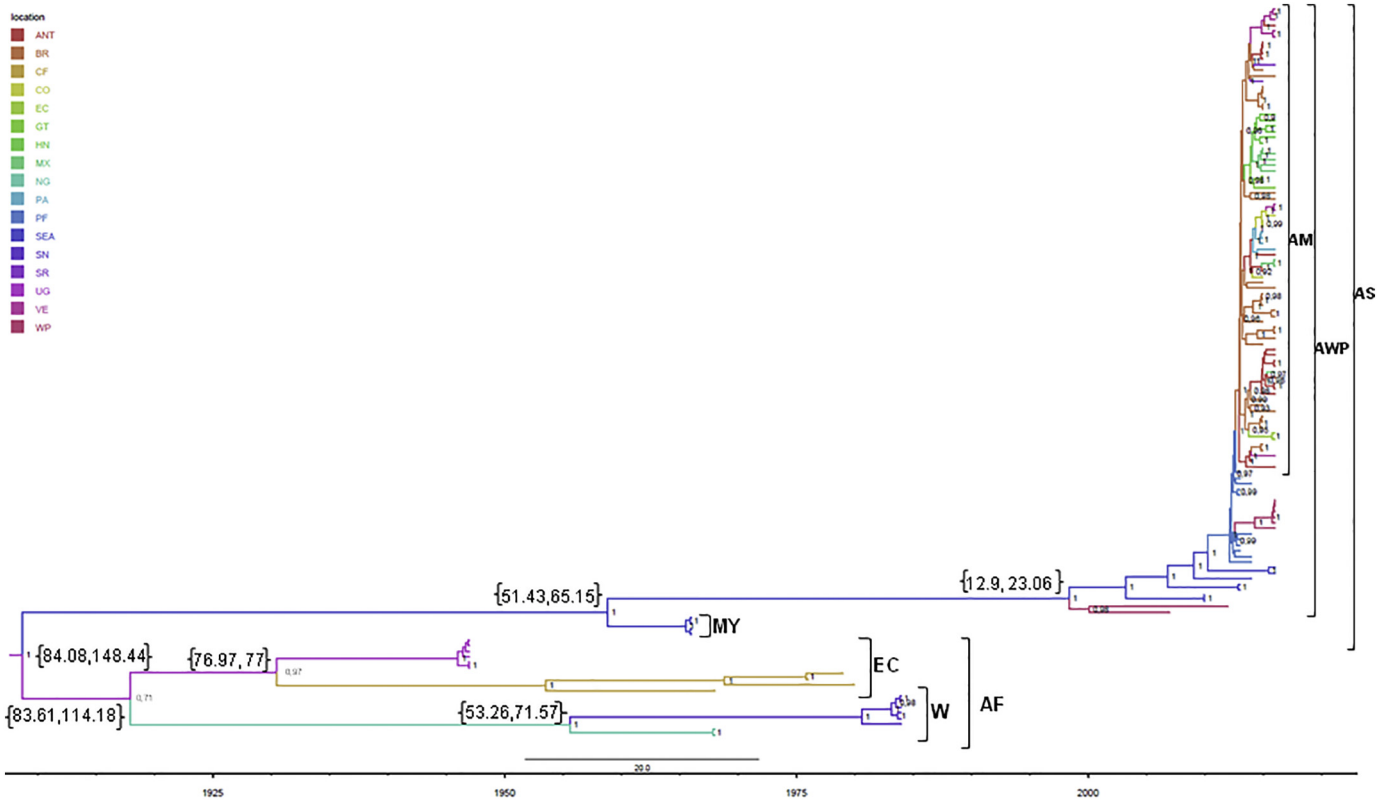

Fig. 2. Phylogeographic analysis of 131 Zika virus isolates in the world. The branches of the maximum clade credibility (MCC) tree are coloured on the basis of the most probable location of the descendent nodes (ANT = Antille; BR = Brazil; *CF* = Central African Republic; CO = Colombia; EC = Ecuador; GT = Guatemala; HN = Honduras; MX = Mexico; NG = Nigeria; PA = Panama; PF = French Polynesia; SEA = South Eastern Asia; SN = Senegal; SR = Suriname; UG = Uganda; VE = Venezuela; WP = Western Pacific). The numbers on the internal nodes indicate posterior probabilities ≥ to 0.7, and the scale at the bottom of the tree represents calendar years. The main geographical clades are highlighted.

Table 3
Posterior probability, estimated times of the most recent common ancestors (tMRCAs) of the main clades and credibility intervals (95% HPD), with calendar years, most probable locations, and state posterior probabilities (spp) of the 131 complete genomes of *Zika virus*.

Nodes	pp[2]	tMRCA[1]			Years			Location	stpp[4]
		Mean	95% HPD[3] Lower	95% HPD[3] Upper	Mean	95% HPD[3] Lower	95% HPD[3] Upper		
Root	1	114,2	148,44	84,08	1902	1868	1932	UG	0,26
African	0,71	98,6	114,18	83,61	1917,4	1901,8	1932,4	UG	0,34
Eastern Central	0,97	86,1	97,77	76	1929,9	1918,2	1940	UG	0,51
Western	1	61,6	71,57	53,26	1954,3	1944,43	1962,7	NG	0,57
Asian	1	57,7	65,15	51,43	1958,2	1950,85	1964,6	SEA	0,81
American	1	3,1	3,46	2,77	2012,9	2012,5	2013,22	BR	0,83

[1] tMRCA: time of the most Recent Common Ancestor.
[2] pp: posterior probability.
[3] HPD: highest posterior density.
[4] stpp: state posterior probability.

available in public databases obtained in over 70 years since the first isolates in 1947 (Giovanetti et al., 2016; Liang et al., 2017), whereas others concentrated the study only on American (Boskova et al., 2018; Faria et al., 2017) or Asian isolates (Pettersson et al., 2018).

In this study we reconstructed the spatiotemporal dynamics of ZIKV at a global and a local scale by using, for the first time, two different phylogeographic approaches: a discrete and a continuous diffusion model.

In order to investigate the phylogenetic information contained in partial genes, we made a root-to-tip regression analysis that showed a sufficient temporal structure (R2 = 0.93) in the whole genome dataset, whereas the results of analyses of the individual gene datasets were ambiguous in terms of locating the best-fitting root position and estimates of the evolutionary rates and root - tMRCA. For these reasons, the dataset was also analysed for the presence of recombination, which revealed four recombinant genomes isolated in Senegal in different years (1968, 1997 and 2001), carrying a total of 12 recombination breakpoints. Homologous recombinations in ZIKV has been previously described (Faye et al., 2014; Han et al., 2016) also in other flaviviruses (Simon-Loriere and Holmes, 2011), and may explain some of the discrepancies in dating the origin of the virus, particularly when this is based on partial genomes.

However, although it therefore seems to be essential to use whole genomes and exclude recombinant sequences in order to obtain

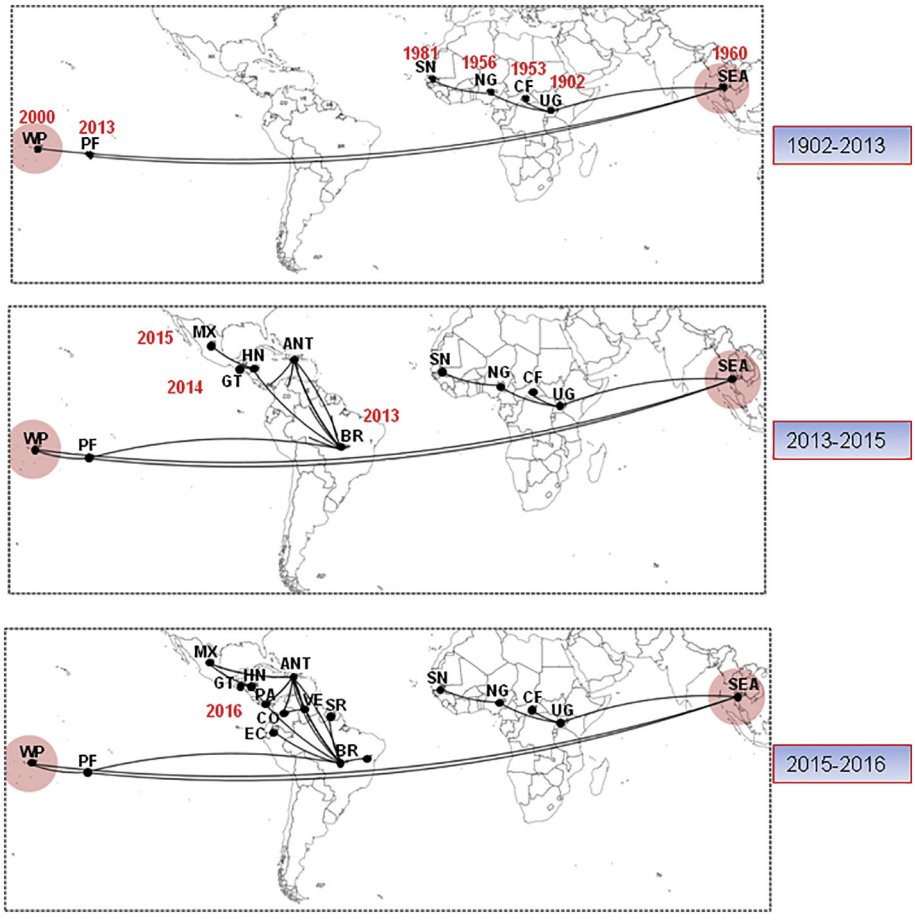

Fig. 3. Spatio-temporal dynamics of the Zika virus epidemic in the world. The figure summarises the most significant migration links in the involved areas.

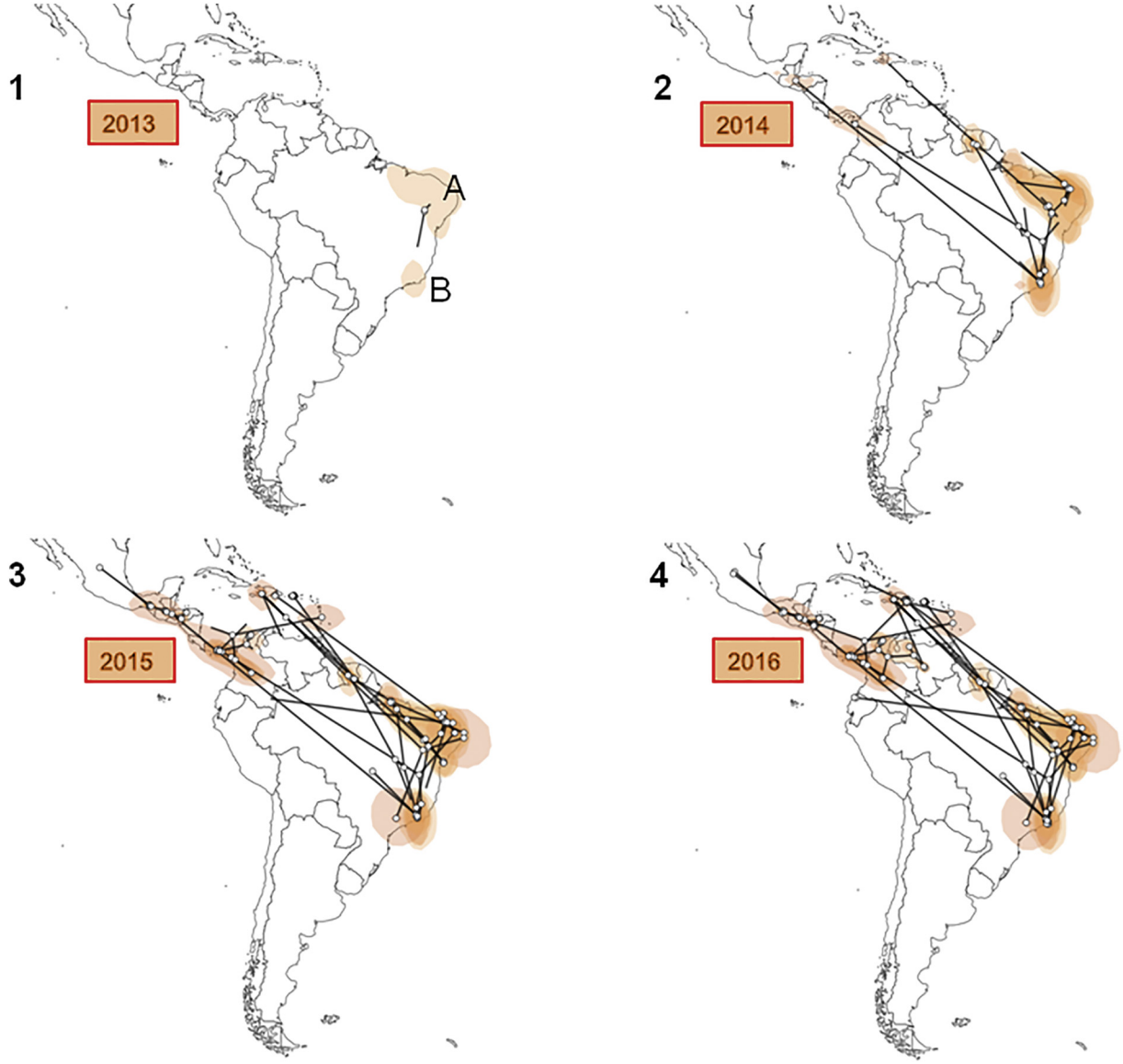

Fig. 4. Spatio-temporal dynamics of the Zika virus epidemic in the Americas. The figure summarises the most significant migration links in the involved area.

unbiased inferences, complete genomes are relatively scarce because of the difficulties in performing extensive genome sequencing (Metsky et al., 2017). The low level of viremia frequently requires preparatory cultural enrichment of the virus, and so we have developed a protocol for the whole genome sequencing of ZIKV that was implemented on the Illumina NGS platform. The full genome of an Italian sample of human semen that has been previously characterised using a Sanger-based method was further analysed using our NGS protocol, revealing over 99% identity with KY003154. Our sequence was aligned with 130 other complete ZIKV genomes for which the sampling dates and locations were known that were retrieved from public databases at the time the analysis began.

Our findings indicate that the mean evolutionary rate of the viral genome was between 8.5 and 22×10^{-4}, similar to the estimates previously obtained by other authors (evolutionary rates ranging from 6 to 13×10^{-4}) (Liang et al., 2017). On the basis of this evolutionary rate, we estimated a tree-root tMRCA (varying from 84 to 148 years), corresponding to the late 19th and early 20th centuries. In agreement with Liang (Liang et al., 2017), who only studied partial genes, our data suggest that the Zika virus emerged > 100 years ago as an endemic/epidemic infection, probably after a period of sylvatic circulation in the rainforests of the east Africa. The virus then spread to West Africa in the 1950s and, a few years later (1958s), to south-east Asia (Malaysia). After a long period of about 40 years without any new isolation, causing a bottleneck effect (highlighted by the long branch of the tree connecting the older Malaysian strains to the new SEA isolates), the virus reappeared in the Pacific islands (Micronesia) in the early 2000s, and caused the first significant outbreak on the Island of Yap in 2007. The virus again left south-east Asia and spread to French Polynesia where it caused the largest outbreak ever recorded before that time. Our estimated tMRCA showed that the virus had been present in French Polynesia at least since the second half of 2012 even if the first recordable cases were in late 2013 (between the 41st week of 2013 and the 14th week of 2014). Our data confirm that ZIKV reached French Polynesia from south-east Asia not from the Island of Yap, as suggested by others (Pettersson et al., 2018; Weaver et al., 2016). Finally, it moved from French Polynesia to the Americas (Brazil) and spread throughout the continent within a few years. Interestingly, the skyline analysis showed an exponential increase in the effective number of infections from early 2013 to late 2014, which corresponds with the spread of the virus in the Western hemisphere, according to our reconstruction (Supplementary Fig. 2).

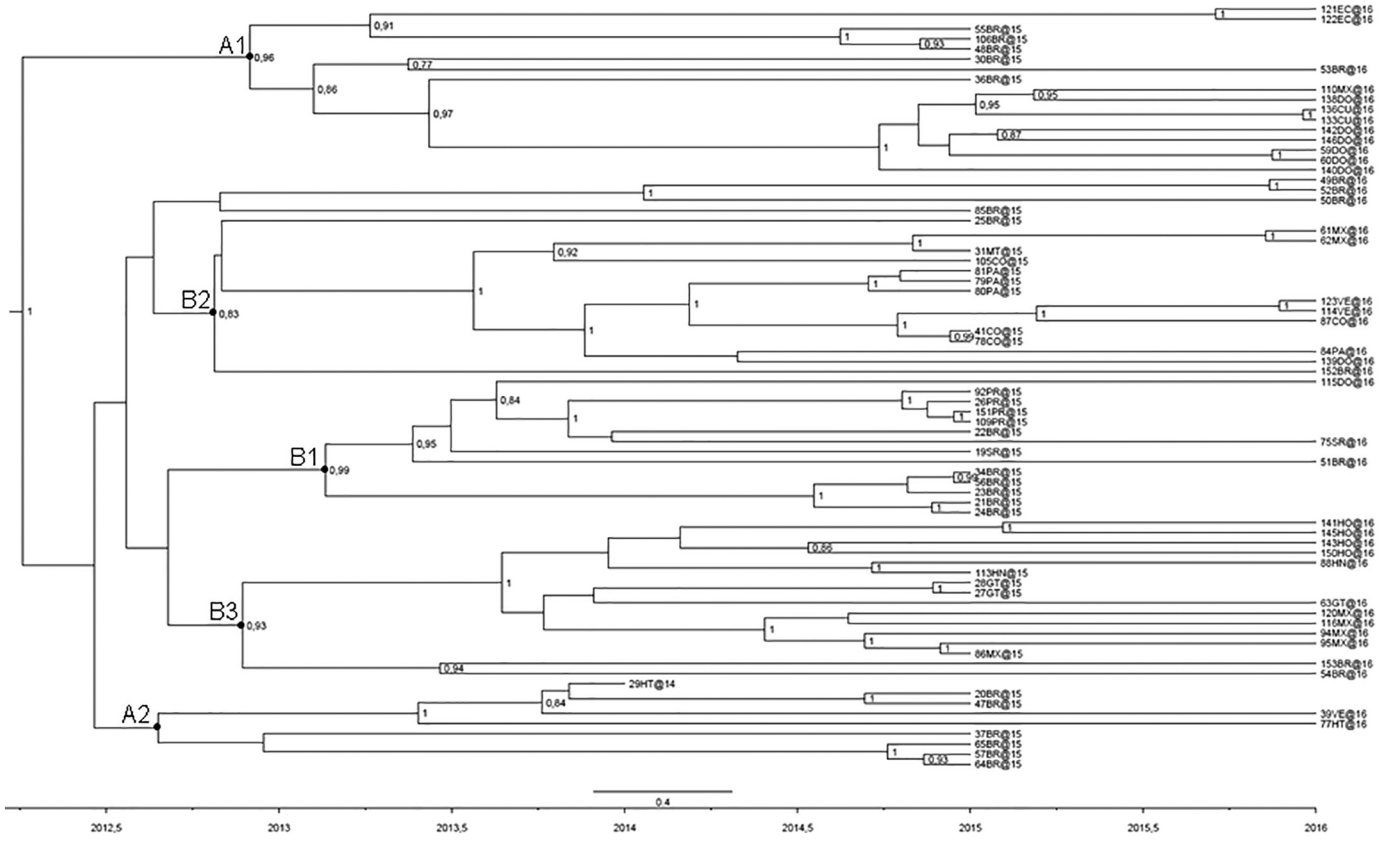

Fig. 5. Phylogeographical analysis of 76 Zika virus isolates in the Americas. The numbers on the internal nodes indicate posterior probabilities of > 0.7, and the scale at the bottom of the tree represents calendar years. The main geographical clades are highlighted.

Table 4
The main clades, calendar months, most probable locations with mean estimated coordinates, and posterior probabilities (pp) of the 76 complete genomes of *Zika virus*.

Clades	pp[a]	Longitude	Latitude	Location	Month	Month L[e]	MONTH U[f]
Root tree	1	41.126	9.532	BRA_NE[b]	October 2012	August 2013	December 2011
A1	0.96	38.8	7.204	BRA_NE[b]	June 2013	May 2014	August 2012
A2	0.12	42.056	8.41	BRA_NE[b]	March 2013	November 2013	May 2012
B2	0.82	42.457	20.003	BRA_SE[c]	April 2013	January 2014	August 2012
B1	0.99	45.678	13.439	BRA_C[d]	August 2013	May 2014	December 2012
B3	0.93	43.377	21.865	BRA_SE[c]	May 2013	February 2014	October 2012

[a] Posterior probability.
[b] North-east Brazil.
[c] South-east Brazil.
[d] Centre Brazil.
[e] Lower.
[f] Upper.

The added value of this study comes from the use of a continuous diffusion model to reconstruct the spread of ZIKV in the Americas. Discrete phylogeography requires the grouping of isolates into categories that can be based on political-administrative boundaries, or other geographically or epidemiologically homogenous areas/populations. However, the grouping is frequently arbitrary and often lacks precision in reconstructing flow rates between different spatial areas, in particular when there are ambiguities in assigning the isolates to one group or another. Continuous phylogeography allows these limitations to be overcome by identifying isolates on the basis of geographical coordinates corresponding to the latitude and longitude of the sampling location. We deduced these coordinates on the basis of the patients' data available, and the use of diffusion models made the reconstruction of the spread of the epidemic in the Americas more precise by allowing ancestral sequences to reside at any location in a continuous bi-dimensional space whereas, in the case of discrete approaches, there is no way to infer the ancestral location from the tree if it is not present in the sampled location set.

Our continuous phylogeographical reconstruction allowed an estimate of the possible geographical coordinates of the entry location and an estimated 80% probability region covering the entire north-east of Brazil, thus suggesting that ZIKV entered Brazil in late 2012. Just two years later, in early 2015, cases of a “dengue-like syndrome”, that probably represented the first cases of the ZIKV epidemic, were reported in two cities: Natal (in the State of Rio Grande del Norte) and Camaçari (in the metropolitan area of Bahia) (Campos et al., 2015; Zanluca et al., 2015). Both cities are included in the 80% probability region of our estimated tree-root even if the analysed samples did not include any coming from these places which underlines the importance of using these models.

Interestingly, the analysis also identified a second high probability region surrounding Rio de Janeiro at the very beginning of the epidemic (2013) as a further initial area of viral dissemination across the Americas. A recent study using a data-driven stochastic and spatial epidemic model considering the period between April 2013 and June 2014 estimated the arrival of ZIKV in the Americas in August 2013 in Rio de Janeiro or north-east Brazil, where mosquito density and DENV transmission is highest (Zhang et al., 2017). This is in line with our spatial reconstruction, although our tree root tMRCA is several months earlier (October 2012, with an upper estimate of August 2013). Furthermore, a recent molecular survey identified the presence of ZIKV RNA in samples collected in March–May 2013 in Tijuca (in the metropolitan area of Rio de Janeiro) from patients with acute febrile syndrome negative for DENV RNA (Passos et al., 2017), and other phylogenetic studies based on human and entomological samples support the presence of ZIKV in Brazil in late 2012 and early 2013 (Ayllon et al., 2017; Metsky et al., 2017).

Two waves of the ZIKV epidemic in Brazil have been hypothesised: an early wave in north-east Brazil and a second in south-east Brazil (Zhang et al., 2017). Our continuous phylogeographical analysis of the American clades identified two clades supporting the hypothesis of a first epidemic wave starting in north-east Brazil in early 2013 that initially spread locally (A1) and reached Haiti (A2). Previous authors have suggested an initial wave of ZIKV in Haiti (Lednicky et al., 2016). The same viral strains were only later exported to Santo Domingo, Cuba and Ecuador.

The second epidemic wave starting in the region surrounding Rio de Janeiro (B) spread over a larger area and give rise to new dispersal centres in Suriname, Panama/Colombia and Honduras/Guatemala from which the infection further spread to the Caribbean islands and Central America. The more extensive spread of the strains originating in region B may have been due to the larger passenger flows related to the importance of Rio as a Brazilian transportation hub (Zhang et al., 2017).

Globally, the virus moved north-west from Brazil to the central-north America at an estimated mean diffusion rate of 760.8 km/year (95% HPD 596–913 km/year) between early 2013 and 2016.

It has recently been suggested that ZIKV may have been imported into the Americas as a consequence of the Confederations Cup held in Brazil in June 2013. It is probable that, during this event, many athletes and supporters from affected areas (French Polynesia) probably travelled through a large area of eastern Brazil stretching from Fortaleza to Rio de Janeiro. An alternative hypothesis is the simultaneous arrival of the same south-east Asian virus in French Polynesia and Brazil (Zhang et al., 2017), but this does not seem to be confirmed by our analysis, which indicates that ZIKV entered Polynesia before reaching Brazil.

One question that deserves to be clarified is the cryptic circulation of the virus in Brazil before the first cases of ZIKV infection were reported in 2013–2015; this also seems to have occurred in Polynesia, where the estimated first entry of the virus is at least one year before the first human cases were reported. The frequency of asymptomatic infections and the presence in the same area of viruses causing infections with similar outcomes (such as *Chikungunya* and *Dengue* viruses) may have masked the initial spread of the virus. Moreover, sylvatic viral circulation can be hypothesised, even if there is only limited evidence of the exposure of non-human primates to Zika virus in the new world (Chiu et al., 2017; Moreira-Soto et al., 2018). Other factors such as the density of the vector population and the seasonality nature of vector abundance (Lana et al., 2014) may also affect the duration of silent circulation.

In the absence of any known animal reservoirs and a possible enzootic circulation such as in the case of *West Nile* Virus, which is known to have undergone a latent period of circulation before the appearance of the first human cases (Zehender et al., 2017), the possible role of transmission routes other than *Aedes* mosquito bites should be investigated. It has been reported that the infection can be sexually transmitted due to its persistent presence in the semen of affected males, even if they are asymptomatic. Recent studies have shown that the sexual transmission may contribute to increase the final size and the persistence of the epidemic (Gao et al., 2016). In particular, a recent study showed that up to 47% of ZIKV cases may be due to sexual contacts when *Aedes* mosquitoes are also present (Towers et al., 2016). Sexual transmission and possibly migration of recently infected subjects (Baca-Carrasco and Velasco-Hernandez, 2016; Olawoyin and Kribs, 2018) may have initially contributed to the slow and hidden circulation of the virus before the accumulation of a critical number of infected humans and mosquitoes.

This is the first study, to our knowledge, that provides an estimate of the geographic origin and diffusion pathways of ZIKV in America across a continuous space. Moreover, it provides a new ZIKV genome obtained through NGS platform.

The main limitation of this study is that it analysed a relatively small number of complete ZIKV genomes, although it must be remembered that the databases included only 135 complete genomes at the time the study was started. It is not a limitation in terms of coalescent theory, which considers small samples of whole populations (Griffiths and Tavare, 1994) but, as it may be a problem for phylogeographical studies in terms of sampled locations, we partially compensated this by using continuous phylogeography. Unfortunately, the scarcity of publicly available sequences is essentially due to the difficulty in detecting the virus in samples and the need for cultures in order to have sufficient material (Metsky et al., 2017).

Our data underline the importance of genome characterisation and the use of phylogeography in the surveillance of emerging infections.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2019.04.006.

Acknowledgements

This work was partially financed by the “NANOMAX” Bandiera project.

(2013–2015) funded by Italian Ministry for Education, University and Research (grant number G42I12000180005) to GZ.

This study was also partially supported by funds from Lombardy Region and grant from.

the Ministero della Salute, Ricerca Corrente Fondazione Istituto Di Ricovero e Cura a.

Carattere Scientifico Policlinico San Matteo, grant no. 80206 to EP.

References

Ayllon, T., Campos, R.M., Brasil, P., Morone, F.C., Camara, D.C.P., Meira, G.L.S., Tannich, E., Yamamoto, K.A., Carvalho, M.S., Pedro, R.S., Schmidt-Chanasit, J., Cadar, D., Ferreira, D.F., Honorio, N.A., 2017. Early evidence for Zika virus circulation among Aedes aegypti mosquitoes, Rio de Janeiro, Brazil. Emerg. Infect. Dis. 23, 1411–1412.

Baca-Carrasco, D., Velasco-Hernandez, J.X., 2016. Sex, mosquitoes and epidemics: an evaluation of Zika disease dynamics. Bull. Math. Biol. 78, 2228–2242.

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M.A., Alekseyenko, A.V., 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol. Biol. Evol. 29, 2157–2167.

Barzon, L., Pacenti, M., Franchin, E., Lavezzo, E., Trevisan, M., Sgarabotto, D., Palu, G., 2016. Infection dynamics in a traveller with persistent shedding of Zika virus RNA in semen for six months after returning from Haiti to Italy, January 2016. Euro Surveill 21, 1560–7917.

Bielejec, F., Baele, G., Vrancken, B., Suchard, M.A., Rambaut, A., Lemey, P., 2016. SpreaD3: interactive visualization of spatiotemporal history and trait evolutionary processes. Mol. Biol. Evol. 33, 2167–2169.

Bloomquist, E.W., Lemey, P., Suchard, M.A., 2010. Three roads diverged? Routes to phylogeographic inference. Trends Ecol. Evol. 25, 626–632.

Boni, M.F., Posada, D., Feldman, M.W., 2007. An Exact Nonparametric Method for Inferring Mosaic Structure in Sequence Triplets. Genetics 176, 1035–1047.

Boskova, V., Stadler, T., Magnus, C., 2018. The influence of phylodynamic model specifications on parameter estimates of the Zika virus epidemic. Virus Evol. 4, vex044.

Buechler, C.R., Bailey, A.L., Weiler, A.M., Barry, G.L., Breitbach, M.E., Stewart, L.M., Jasinska, A.J., Freimer, N.B., Apetrei, C., Phillips-Conroy, J.E., Jolly, C.J., Rogers, J., Friedrich, T.C., O'Connor, D.H., 2017. Seroprevalence of Zika virus in wild African green monkeys and baboons. mSphere 2, 00392–00316.

Calvet, G.A., Filippis, A.M., Mendonca, M.C., Sequeira, P.C., Siqueira, A.M., Veloso, V.G., Nogueira, R.M., Brasil, P., 2016. First detection of autochthonous Zika virus transmission in a HIV-infected patient in Rio de Janeiro, Brazil. J. Clin. Virol. 74, 1–3.

Calvez, E., Mousson, L., Vazeille, M., O'Connor, O., Cao-Lormeau, V.M., Mathieu-Daude,

F., Pocquet, N., Failloux, A.B., Dupont-Rouzeyrol, M., 2018. Zika virus outbreak in the Pacific: vector competence of regional vectors. PLoS Negl. Trop. Dis. 12, e0006637.
Campos, G.S., Bandeira, A.C., Sardi, S.I., 2015. Zika virus outbreak, Bahia, Brazil. Emerg. Infect. Dis. 21, 1885–1886.
Chiu, C.Y., Sanchez-San Martin, C., Bouquet, J., Li, T., Yagi, S., Tamhankar, M., Hodara, V.L., Parodi, L.M., Somasekar, S., Yu, G., Giavedoni, L.D., Tardif, S., Patterson, J., 2017. Experimental Zika virus inoculation in a New World monkey model reproduces key features of the human infection. Sci. Rep. 7, 17126.
De Smet, B., Van den Bossche, D., van de Werve, C., Mairesse, J., Schmidt-Chanasit, J., Michiels, J., Arien, K.K., Van Esbroeck, M., Cnops, L., 2016. Confirmed Zika virus infection in a Belgian traveler returning from Guatemala, and the diagnostic challenges of imported cases into Europe. J. Clin. Virol. 80, 8–11.
Djikeng, A., Halpin, R., Kuzmickas, R., Depasse, J., Feldblyum, J., Sengamalay, N., Afonso, C., Zhang, X., Anderson, N.G., Ghedin, E., Spiro, D.J., 2008. Viral genome sequencing by random priming methods. BMC Genomics 9, 5.
D'Ortenzio, E., Matheron, S., Yazdanpanah, Y., de Lamballerie, X., Hubert, B., Piorkowski, G., Maquart, M., Descamps, D., Damond, F., Leparc-Goffart, I., 2016. Evidence of sexual transmission of Zika virus. N. Engl. J. Med. 374, 2195–2198.
Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22, 1185–1192.
Duffy, M.R., Chen, T.H., Hancock, W.T., Powers, A.M., Kool, J.L., Lanciotti, R.S., Pretrick, M., Marfel, M., Holzbauer, S., Dubray, C., Guillaumot, L., Griggs, A., Bel, M., Lambert, A.J., Laven, J., Kosoy, O., Panella, A., Biggerstaff, B.J., Fischer, M., Hayes, E.B., 2009. Zika virus outbreak on Yap Island, Federated States of Micronesia. N. Engl. J. Med. 360, 2536–2543.
Faria, N.R., Quick, J., Claro, I.M., Theze, J., de Jesus, J.G., Giovanetti, M., Kraemer, M.U.G., Hill, S.C., Black, A., da Costa, A.C., Franco, L.C., Silva, S.P., Wu, C.H., Raghwani, J., Cauchemez, S., du Plessis, L., Verotti, M.P., de Oliveira, W.K., Carmo, E.H., Coelho, G.E., Santelli, A., Vinhal, L.C., Henriques, C.M., Simpson, J.T., Loose, M., Andersen, K.G., Grubaugh, N.D., Somasekar, S., Chiu, C.Y., Munoz-Medina, J.E., Gonzalez-Bonilla, C.R., Arias, C.F., Lewis-Ximenez, L.L., Baylis, S.A., Chieppe, A.O., Aguiar, S.F., Fernandes, C.A., Lemos, P.S., Nascimento, B.L.S., Monteiro, H.A.O., Siqueira, I.C., de Queiroz, M.G., de Souza, T.R., Bezerra, J.F., Lemos, M.R., Pereira, G.F., Loudal, D., Moura, L.C., Dhalia, R., Franca, R.F., Magalhaes, T., Marques Jr., E.T., Jaenisch, T., Wallau, G.L., de Lima, M.C., Nascimento, V., de Cerqueira, E.M., de Lima, M.M., Mascarenhas, D.L., Neto, J.P.M., Levin, A.S., Tozetto-Mendoza, T.R., Fonseca, S.N., Mendes-Correa, M.C., Milagres, F.P., Segurado, A., Holmes, E.C., Rambaut, A., Bedford, T., Nunes, M.R.T., Sabino, E.C., Alcantara, L.C.J., Loman, N.J., Pybus, O.G., 2017. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature 546, 406–410.
Faye, O., Freire, C.C., Iamarino, A., Faye, O., de Oliveira, J.V., Diallo, M., Zanotto, P.M., Sall, A.A., 2014. Molecular evolution of Zika virus during its emergence in the 20(th) century. PLoS Negl. Trop. Dis. 8, e2636.
Gao, D., Lou, Y., He, D., Porco, T.C., Kuang, Y., Chowell, G., Ruan, S., 2016. Prevention and control of Zika as a mosquito-borne and sexually transmitted disease: a mathematical Modeling analysis. Sci. Rep. 6, 28070.
Garcia-Luna, S.M., Weger-Lucarelli, J., Ruckert, C., Murrieta, R.A., Young, M.C., Byas, A.D., Fauver, J.R., Perera, R., Flores-Suarez, A.E., Ponce-Garcia, G., Rodriguez, A.D., Ebel, G.D., Black, W.C.t., 2018. Variation in competence for ZIKV transmission by Aedes aegypti and Aedes albopictus in Mexico. PLoS Negl. Trop. Dis. 12, e0006599.
Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2000. Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16 (7), 573–582.
Giovanetti, M., Milano, T., Alcantara, L.C., Carcangiu, L., Cella, E., Lai, A., Lo Presti, A., Pascarella, S., Zehender, G., Angeletti, S., Ciccozzi, M., 2016. Zika virus spreading in South America: evolutionary analysis of emerging neutralizing resistant Phe279Ser strains. Asian Pac J Trop Med 9, 445–452.
Griffiths, R.C., Tavare, S., 1994. Sampling theory for neutral alleles in a varying environment. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 344, 403–410.
Gubler, D.J., Vasilakis, N., Musso, D., 2017. History and emergence of Zika virus. J. Infect. Dis. 216, S860–S867.
Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.
Han, J.F., Jiang, T., Ye, Q., Li, X.F., Liu, Z.Y., Qin, C.F., 2016. Homologous recombination of Zika viruses in the Americas. J. Inf. Secur. 73, 87–88.
Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., Gibson, T.J., 1998. Multiple sequence alignment with Clustal X. Trends Biochem. Sci. 23, 403–405.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.
Lana, R.M., Carneiro, T.G., Honorio, N.A., Codeco, C.T., 2014. Seasonal and nonseasonal dynamics of Aedes aegypti in Rio de Janeiro, Brazil: fitting mathematical models to trap data. Acta Trop. 129, 25–32.
Kosakovsky, S.L., Posada, D., Gravenor, M.B, Woelk, C.H, Frost, S.D.W., 2006. GARD: a genetic algorithm for recombination detection. Bioinforma. Appl. Note 22 (24), 3096–3098.
Lazear, H.M., Diamond, M.S., 2016. Zika virus: new clinical syndromes and its emergence in the Western hemisphere. J. Virol. 90, 4864–4875.
Lednicky, J., Beau De Rochars, V.M., El Badry, M., Loeb, J., Telisma, T., Chavannes, S., Anilis, G., Cella, E., Ciccozzi, M., Rashid, M., Okech, B., Salemi, M., Morris, J.G., Jr., 2016. Zika virus outbreak in Haiti in 2014: molecular and clinical data. PLoS Negl. Trop. Dis. 10, e0004687.
Lemey, P., Rambaut, A., Drummond, A.J., Suchard, M.A., 2009. Bayesian phylogeography finds its roots. PLoS Comput. Biol. 5, e1000520.
Lemey, P., Rambaut, A., Welch, J.J., Suchard, M.A., 2010. Phylogeography takes a relaxed random walk in continuous space and time. Mol. Biol. Evol. 27, 1877–1885.
Liang, D., Leung, R.K.K., Lee, S.S., Kam, K.M., 2017. Insights into intercontinental spread of Zika virus. PLoS One 12, e0176710.
Marchette, N.J., Garcia, R., Rudnick, A., 1969. Isolation of Zika virus from Aedes aegypti mosquitoes in Malaysia. Am. J. Trop. Med. Hyg. 18, 411–415.
Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Vir.Evol. 1.
Martin, D.P., Williamson, C., Posada, D., 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics Applications Note Vol.21. pp. 260–262 2.
Metsky, H.C., Matranga, C.B., Wohl, S., Schaffner, S.F., Freije, C.A., Winnicki, S.M., West, K., Qu, J., Baniecki, M.L., Gladden-Young, A., Lin, A.E., Tomkins-Tinch, C.H., Ye, S.H., Park, D.J., Luo, C.Y., Barnes, K.G., Shah, R.R., Chak, B., Barbosa-Lima, G., Delatorre, E., Vieira, Y.R., Paul, L.M., Tan, A.L., Barcellona, C.M., Porcelli, M.C., Vasquez, C., Cannons, A.C., Cone, M.R., Hogan, K.N., Kopp, E.W., Anzinger, J.J., Garcia, K.F., Parham, L.A., Ramirez, R.M.G., Montoya, M.C.M., Rojas, D.P., Brown, C.M., Hennigan, S., Sabina, B., Scotland, S., Gangavarapu, K., Grubaugh, N.D., Oliveira, G., Robles-Sikisaka, R., Rambaut, A., Gehrke, L., Smole, S., Halloran, M.E., Villar, L., Mattar, S., Lorenzana, I., Cerbino-Neto, J., Valim, C., Degrave, W., Bozza, P.T., Gnirke, A., Andersen, K.G., Isern, S., Michael, S.F., Bozza, F.A., Souza, T.M.L., Bosch, I., Yozwiak, N.L., MacInnis, B.L., Sabeti, P.C., 2017. Zika virus evolution and spread in the Americas. Nature 546, 411–415.
Mittal, R., Nguyen, D., Debs, L.H., Patel, A.P., Liu, G., Jhaveri, V.M., SI, S.K., Mittal, J., Bandstra, E.S., Younis, R.T., Chapagain, P., Jayaweera, D.T., Liu, X.Z., 2017. Zika virus: an emerging Global Health threat. Front. Cell. Infect. Microbiol. 7, 486.
Moreira-Soto, A., Carneiro, I.O., Fischer, C., Feldmann, M., Kummerer, B.M., Silva, N.S., Santos, U.G., Souza, B., Liborio, F.A., Valenca-Montenegro, M.M., Laroque, P.O., da Fontoura, F.R., Oliveira, A.V.D., Drosten, C., de Lamballerie, X., Franke, C.R., Drexler, J.F., 2018. Limited evidence for infection of urban and Peri-urban nonhuman primates with Zika and Chikungunya viruses in Brazil. mSphere 3 00523–00517.
Musso, D., Bossin, H., Mallet, H.P., Besnard, M., Broult, J., Baudouin, L., Levi, J.E., Sabino, E.C., Ghawche, F., Lanteri, M.C., Baud, D., 2018. Zika virus in French Polynesia 2013-14: anatomy of a completed outbreak. Lancet Infect. Dis. 18, e172–e182.
Olawoyin, O., Kribs, C., 2018. Effects of multiple transmission pathways on Zika dynamics. Infect Dis Model 3, 331–344.
Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible Emergence of New Geminiviruses by Frequent Recombination. Virol. 265, 218–225.
Passos, S.R.L., Borges Dos Santos, M.A., Cerbino-Neto, J., Buonora, S.N., Souza, T.M.L., de Oliveira, R.V.C., Vizzoni, A., Barbosa-Lima, G., Vieira, Y.R., Silva de Lima, M., Hokerberg, Y.H.M., 2017. Detection of Zika virus in April 2013 patient samples, Rio de Janeiro, Brazil. Emerg. Infect. Dis. 23, 2120–2121.
Pettersson, J.H., Bohlin, J., Dupont-Rouzeyrol, M., Brynildsrud, O.B., Alfsnes, K., Cao-Lormeau, V.M., Gaunt, M.W., Falconar, A.K., de Lamballerie, X., Eldholm, V., Musso, D., Gould, E.A., 2018. Re-visiting the evolution, dispersal and epidemiology of Zika virus in Asia. Emerg. Microbes Infect. 7, 79.
Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.
Posada, D., Krandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. PNAS 98 (24), 13757–13762.
Rambaut, A., Lam, T.T., Max Carvalho, L., Pybus, O.G., 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly path-O-gen). Virus Evol. 2, vew007.
Sebastian, U.U., Ricardo, A.V.A., Alvarez, B.C., Cubides, A., Luna, A.F., Arroyo-Parejo, M., Acuna, C.E., Quintero, A.V., Villareal, O.C., Pinillos, O.S., Vieda, E., Bello, M., Pena, S., Duenas-Castell, C., Rodriguez, G.M.V., Ranero, J.L.M., Lopez, R.L.M., Olaya, S.G., Vergara, J.C., Tandazo, A., Ospina, J.P.S., Leyton Soto, I.M., Fowler, R.A., Marshall, J.C., 2017. Zika virus-induced neurological critical illness in Latin America: severe Guillain-Barre syndrome and encephalitis. J. Crit. Care 42, 275–281.
Simonin, Y., van Riel, D., Van de Perre, P., Rockx, B., Salinas, S., 2017. Differential virulence between Asian and African lineages of Zika virus. PLoS Negl. Trop. Dis. 11, e0005821.
Simon-Loriere, E., Holmes, E.C., 2011. Why do RNA viruses recombine? Nat. Rev. Microbiol. 9, 617–626.
Smith, J.M., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.
Strimmer, K., von Haeseler, A., 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl. Acad. Sci. U. S. A. 94, 6815–6819.
Towers, S., Brauer, F., Castillo-Chavez, C., Falconar, A.K.I., Mubayi, A., Romero-Vivas, C.M.E., 2016. Estimate of the reproduction number of the 2015 Zika virus outbreak in Barranquilla, Colombia, and estimation of the relative role of sexual transmission. Epidemics 17, 50–55.
Weaver, S.C., Costa, F., Garcia-Blanco, M.A., Ko, A.I., Ribeiro, G.S., Saade, G., Shi, P.Y., Vasilakis, N., 2016. Zika virus: history, emergence, biology, and prospects for control. Antivir. Res. 130, 69–80.
Zanluca, C., Melo, V.C., Mosimann, A.L., Santos, G.I., Santos, C.N., Luz, K., 2015. First report of autochthonous transmission of Zika virus in Brazil. Mem. Inst. Oswaldo Cruz 110, 569–572.
Zehender, G., Veo, C., Ebranati, E., Carta, V., Rovida, F., Percivalle, E., Moreno, A., Lelli, D., Calzolari, M., Lavazza, A., Chiapponi, C., Baioni, L., Capelli, G., Ravagnan, S., Da Rold, G., Lavezzo, E., Palu, G., Baldanti, F., Barzon, L., Galli, M., 2017. Reconstructing the recent West Nile virus lineage 2 epidemic in Europe and Italy using discrete and continuous phylogeography. PLoS One 12, e0179679.
Zhang, Q., Sun, K., Chinazzi, M., Pastore, Y.P.A., Dean, N.E., Rojas, D.P., Merler, S., Mistry, D., Poletti, P., Rossi, L., Bray, M., Halloran, M.E., Longini Jr., I.M., Vespignani, A., 2017. Spread of Zika virus in the Americas. Proc. Natl. Acad. Sci. U. S. A. 114, E4334–E4343.