Zika virus in the Americas: Early epidemiological and genetic findings


Version Post-print/accepted manuscript

Citation


Publisher’s statement

This is the author’s version of the work. It is posted here by permission of the AAAS for personal use, not for redistribution. The definitive version was published in Science on April 15 2016 in Volume 352, DOI: 10.1126/science.aaf5036.

How to cite TSpace items

Always cite the published version, so the author(s) will receive recognition through services that track citation counts, e.g. Scopus. If you need to cite the page number of the author manuscript from TSpace because you cannot access the published version, then cite the TSpace version in addition to the published version using the permanent URI (handle) found on the record page.

This article was made openly accessible by U of T Faculty. Please tell us how this access benefits you. Your story matters.
Zika virus in the Americas: Early epidemiological and genetic findings

A full list of authors and affiliations appears at the end of the article.

Brazil has experienced an unprecedented epidemic of Zika virus (ZIKV), with ~30,000 cases reported to date. ZIKV was first detected in Brazil in May 2015 and cases of microcephaly potentially associated with ZIKV infection were identified in November 2015. Using next generation sequencing we generated seven Brazilian ZIKV genomes, sampled from four selflimited cases, one blood donor, one fatal adult case, and one newborn with microcephaly and congenital malformations. Phylogenetic and molecular clock analyses show a single introduction of ZIKV into the Americas, estimated to have occurred between May-Dec 2013, more than 12 months prior to the detection of ZIKV in Brazil. The estimated date of origin coincides with an increase in air passengers to Brazil from ZIKV endemic areas, and with reported outbreaks in Pacific Islands. ZIKV genomes from Brazil are phylogenetically interspersed with those from other South American and Caribbean countries. Mapping mutations onto existing structural models revealed the context of viral amino acid changes present in the outbreak lineage; however no shared amino acid changes were found among the three currently available virus genomes from microcephaly cases. Municipality-level incidence data indicate that reports of suspected microcephaly in Brazil best correlate with ZIKV incidence around week 17 of pregnancy, although this correlation does not demonstrate causation. Our genetic description and analysis of ZIKV isolates in Brazil provide a baseline for future studies of the evolution and molecular epidemiology in the Americas of this emerging virus.

Zika virus (ZIKV) is a single stranded, positive-sense RNA virus with a 10.7 kb genome encoding a single polyprotein that is cleaved into three structural proteins (C, prM/M, E) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) (1). ZIKV is a member of the family *Flaviviridae*, genus *Flavivirus*, and is transmitted among humans by *Aedes* mosquito species such as *A. aegypti*, *A. albopictus*, and *A. africanus*. The virus was first isolated in 1947 from a sentinel rhesus monkey in the Zika forest in Uganda (2) and is classified by sequence analysis into two genotypes, African and Asian (3). In humans, ZIKV infection typically causes a mild and self-limiting illness known as Zika fever (4) accompanied by maculopapular rash, headache, conjunctivitis and myalgia. In April 2007, a large epidemic of Asian genotype ZIKV was reported in Yap Island and Guam, Micronesia (5, 6). Between 2013–2014 the Asian genotype caused epidemics reported in several Pacific Islands, including French Polynesia (7), New Caledonia (8), Cook Islands (9), Tahiti (10), and Easter Island (11).

‡ Corresponding author. oliver.pybus@zoo.ox.ac.uk (O.G.P.); marcionunesbrasil@yahoo.com.br (M.R.T.N.); pedrovasconcelos@iec.pa.gov.br (P.F.C.V.).
By May 2015, ZIKV was reported in Brazil (12) and subsequently in several countries of South and Central America, and the Caribbean. In Brazil nearly 30,000 cases of ZIKV infection had been notified by 30th Jan 2016 (supplementary materials section 1.4). Reported cases in Brazil indicate an epidemic peak in mid-July 2015 (Fig. 1A) and most Brazilian ZIKV cases (93%) were reported in Bahia state (Fig. 1B). ZIKV surveillance in Brazil began after the first reported Brazilian case and is conducted through the national Notifiable Diseases Information System (SINAN), which currently relies on passive case detection and reporting and therefore underestimates incidence (13). ZIKV is now widespread in Brazil, with autochthonous transmission and high incidence notified in 22 out of 27 administrative states (14). ZIKV infection during pregnancy has been hypothesized to cause microcephaly and congenital abnormalities (15–20). The detection of ZIKV in fetal brain tissue (17, 20) and amniotic fluid (21) supports the hypothesis that the virus is transmitted from mother-to-child (22) and the virus infects neural progenitor cells in vitro (23). In Brazil, between Nov 2015 and 30th Jan 2016, 4783 suspected cases of microcephaly were reported electronically to the RESP database (www.resp.saude.gov.br; Ministry of Health, Brazil; see supplementary materials section 1.4) (Fig. 1C), although most suspected cases are still under investigation and a substantial proportion may represent misdiagnosis and over-reporting (24). Using the WHO guidelines for microcephaly diagnosis provided on the 4th March 2016 (25), we identified a total of 1118 suspected microcephaly cases suitable for analysis. The relationship between total per capita ZIKV incidence (Fig. 1B) and per capita suspected microcephaly cases (Fig. 1C) in each state is weak and only significant under non-parametric correlation (p < 0.01) (fig. S1A); noise and uncertainty probably affect both variables. However, the relation is strengthened if suspected microcephaly cases are measured per pregnancy (fig. S1B). For municipalities with reported ZIKV incidence and cases of suspected microcephaly, we used a simple linear model to link microcephaly cases as a function of past ZIKV incidence (supplementary materials section 1.5). Suspected microcephaly cases are best predicted by ZIKV incidence during week 17 of pregnancy on average (95% confidence interval of mean = +/-0.11 weeks), or week 14 for suspected severe microcephaly cases (+/-0.08 weeks), in general agreement with individual reports of the timing of ZIKV symptoms in mothers of infants with microcephaly (16, 19, 21). We stress that these results quantify only the correlation between ZIKV and suspected microcephaly and does not demonstrate a causal link. Work is ongoing to establish whether or not ZIKV is a causal factor in microcephaly and other conditions (15–17, 23, 26).

We used phylogenetic, epidemiological, and mobility data to quantify ZIKV evolution and explore the introduction of the virus to the Americas. As part of ongoing surveillance by the Brazilian Ministry of Health, national laboratories, and other institutions, we used next generation sequencing to generate seven complete ZIKV coding region sequences from samples collected during the outbreak, including one from a deceased newborn with microcephaly and congenital malformations collected in Ceará and one from a fatal adult case with lupus and rheumatoid disease from Maranhão State (Fig. 1B). None of the Brazilian patients reported overseas travel (information unavailable in one case) and one subject was a blood donor (supplementary materials section 2). A comparison of our genomes with other available Brazilian strains reveals that Brazilian ZIKV isolates differ at multiple nucleotide sites across the 10.3kb coding region. The ZIKV genome recovered
from isolate ZIKSP, from São Paulo, had 32 nucleotide changes compared to the microcephaly case (BeH823339) and 34 to the fatal case from Maranhão (BeH818305). Isolates BeH819966 from Belém, BeH815744 from Paraíba, and BeH18995, from Belém had a maximum of 5 nucleotide changes.

Maximum likelihood analysis of complete coding regions from our and other ZIKV genome sequences reveals that all viruses sampled in the Americas, including those from Brazil, form a robust monophyletic cluster (bootstrap score = 94%) within the Asian genotype (Fig. 2 and fig. S2) and share a common ancestor with the ZIKV strain that circulated in French Polynesia in November 2013 (Fig. 3). Previous analyses of outbreaks of related flaviviruses [e.g., (27, 28)] suggest that, to be informative, molecular epidemiological studies of the current ZIKV epidemic should use full or near-complete coding region sequences.

We used a phylogenetic molecular clock approach to further explore the molecular epidemiology of ZIKV in the Americas. A strong correlation between genetic divergence and sampling time within the outbreak lineage (Fig. 2, inset) shows this approach is appropriate provided that whole genomes are used. The estimated time-scaled phylogeny (Fig. 3A) again contains a well-supported clade of American ZIKV strains (denoted B; posterior probability, PP = 1.00) that share a common ancestor (denoted A) with the French Polynesia lineage (PP = 0.92). Within the American ZIKV lineage (clade B), Brazilian isolates are interspersed among isolates from elsewhere in the Americas. The mingling of ZIKV genomes from different countries reveals ZIKV movement within the Americas since its introduction to the continent. Two observations suggest that the common ancestor of the American ZIKV lineage existed in Brazil. First, Brazil was the first country in the Americas to detect ZIKV (29) and second, Brazilian strains are phylogenetically more diverse within clade B than those from elsewhere. However, these observations may reflect differences in surveillance intensity among countries and more data are required before we can exclude the scenario that ZIKV was introduced to Brazil multiple times from other locations. Although two of three ZIKV-associated microcephaly isolates group together in the phylogeny, there is no reason to posit that this lineage is associated with increased disease severity.

Estimated rates of ZIKV molecular evolution are consistent among different evolutionary models and vary from $0.98 \times 10^{-3}$ to $1.06 \times 10^{-3}$ nucleotide substitutions per site per year (table S3). Although this rate is high compared to whole genome rates for other flaviviruses [e.g., (28)], it is consistent with retrospective analyses of previous epidemics, which show that evolutionary rate estimates decline as the epidemic progresses (30, 31). Hence, this result should not be interpreted as implying that ZIKV in the Americas is unusually mutable. We estimate that the date of the most recent common ancestor (TMRCA) of all Brazilian genomes (clade B) is Aug 2013 to Apr 2014 (95% Bayesian credible intervals, BCIs; point estimate = mid Dec 2013; Fig. 3B). The common ancestor of the French Polynesian and America lineages (clade A) was dated to Dec 2012 to Sep 2013 (BCIs; point estimate = late May 2013; Fig. 3B). The posterior distribution for the age of clade B encompasses the recorded duration of the ZIKV outbreak in 3 of 5 island groups of French Polynesia (4) (Fig. 3C). Divergence date estimates are robust among different combinations of prior distributions, molecular clock models, and coalescent models (supplementary materials).
sections 4 and 5), and are more likely to shift into the past than toward the present as virus genomes accumulate through time (30).

To explore possible routes of entry of ZIKV in Brazil, we collated airline flight data from all countries with reported ZIKV outbreaks between 2012 and end of 2014. From late 2012 we find an increase in the number of travellers arriving in Brazil from these countries, rising from 3775 passengers per month in early 2013 to 5754 passengers per month a year later (Fig. 3C). This increase in visitors to Brazil from ZIKV-affected countries coincides with the period during which ZIKV is estimated to have entered the Americas (i.e., between the TMRCAs of clades A and B) (Fig. 3B and supplementary materials section 5). If the ZIKV epidemic in Brazil did indeed arise from a single introduction then the virus must have circulated in the country for at least 12 months prior to the first case being reported in May 2015. ZIKV clinical symptoms may be confused with those caused by dengue and chikungunya viruses, two endemic and epidemic viruses that co-circulate and share mosquito vectors with ZIKV in Brazil (27, 32, 33). Reliable differential diagnosis is possible only by using improved surveillance and laboratory diagnostics, which are now being implemented throughout the country.

There are two published hypotheses for how ZIKV came to be introduced into Brazil, during (i) the 2014 World Cup soccer tournament (Jun 12th - Jul 13th) (29) or (ii) the Va’a canoe event held in Rio de Janeiro between 12-17 Aug 2014 (34). Alternatively, introduction could have occurred during (iii) the 2013 Confederations Cup soccer tournament (15th–30th Jun 2013). Events (ii) and (iii) notably included competitors from French Polynesia. Our results suggest that the introduction of ZIKV to the Americas predated events (i) and (ii). Although the molecular clock dates are more consistent with the Confederations cup, that event ended before ZIKV cases were first reported in French Polynesia (4). Consequently, we believe that large-scale patterns in human mobility will provide more useful and testable hypotheses about viral introduction and emergence (33, 35, 36) than ad hoc hypotheses focused on specific events.

The ZIKV genome we obtained from a microcephaly case in Ceará Brazil contains eight amino acid changes not observed in any other complete genome in our dataset. However, none of these mutations are shared with either of two recently published genomes from microcephaly cases (16, 21). Thus, if a causal link between Asian lineage ZIKV and microcephaly is confirmed, it is possible that putative viral genetic determinants of disease will be found among the amino acid changes that occur on the ZIKV phylogeny branches ancestral to the French Polynesian and American ZIKV lineages (i.e., the two lineages associated with reports of microcephaly, Guillain-Barré syndrome and congenital abnormalities) (37). Phylogenetic character mapping using parsimony reveals 11 amino acid changes on the four internal branches (labeled with asterisks in Fig. 2; fig. S3) leading to these two lineages. We identified the structures of homologous proteins most closely related to ZIKV proteins (supplementary materials section 7) and used them to map 7 of the 11 amino acid changes in a structural context, to five proteins: the pr-peptide region of prM [changes Val123→Ala123 (V123A) and S139N (S, Ser; N, Asn)], NS1 (A982V), the RNA helicase [NS3; N1902H and Y2086H (H, His; Y, Tyr)], the FtsJ-like methyl transferase domain [NS5; M2634V (M, Met)], and the thumb domain of RNA-directed RNA
polymerase (NS5; M3392V) (fig. S7). None of these mutations are predicted to substantially affect the physicochemical properties of the protein environment, except possibly Y2086H (in the helicase; Fig. S8), which may increase the hydrophilicity of the region. The remaining four amino acid changes could not be accurately mapped due to the absence of suitable related X-ray structures (supplementary materials section 7). Notably, none of the observed changes map to the E glycoprotein ectodomain, the primary target of humoral immune responses against flaviviruses (38, 39). Factors other than viral genetic differences may be important for the proposed pathogenesis of ZIKV; hypothesized factors include co-infection with chikungunya virus (40), previous infection with dengue virus (41), or differences in human genetic predisposition to disease.

Besides vector-borne and mother-to-child transmission, Zika virus may also spread via sexual contact (42, 43) and blood transfusion (44). The evidence of ZIKV in blood donors raises the possibility of ZIKV transmission through transfusion and indicates that it may be prudent to consider the screening of blood donors.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Authors**

Nuno Rodrigues Faria#1,2, Raimundu do Socorro da Silva Azevedo#3, Moritz U.G. Kraemer2, Renato Souza4, Mariana Sequetin Cunha4, Sarah C. Hill2, Julien Thézé2, Michael B. Bonsall2, Thomas A. Bowden5, Ilona Rissanen5, Iray Maria Rocco4, Juliana Silva Nogueira4, Adriana Yurika Maeda4, Fernanda Giseli da Silva Vasami4, Fernando Luiz de Lima Macedo4, Akemi Suzuki4, Sueli Guerreiro Rodrigues3, Ana Cecilia Ribeiro Cruz2, Bruno Tardeli Nunes3, Daniele Barbosa de Almeida Medeiros3, Daniela Sueli Guerreiro Rodrigues3, Alice Louize Nunes Queiroz2, Eliana Vieira Pinto da Silva3, Daniele Freitas Henriques3, Elisabeth Salbe Travassos da Rosa3, Consuelo Silva de Oliveira2, Livia Caricio Martins3, Helena Baldez Vasconcelos2, Livia Medeiros Neves Casseb3, Darlene de Brito Simith3, Jane P. Messina2,6, Leandro Abade2, José Lourenço2, Luiz Carlos Junior Alcantara7, Maricélia Maia de Lima8, Marta Giovannetti7, Simon I. Hay8,5, Rodrigo Santos de Oliveira1, Poliana da Silva Lemos1, Layanna Freitas de Oliveira1, Clayton Pereira Silva de Lima1, Sandro Patroca da Silva1, Janaina Mota de Vasconcelos1, Luciano Franco1, Jedson Ferreira Cardoso1, João Lídio da Silva Gonçalves Viane-Júnior1, Dalana Mir10, Gonzalo Bello10, Edson Delatorre10, Kamran Khan11,12, Marisa Creatore13, Giovannini Evelin Coelho14, Wanderson Kleber de Oliveira14, Robert Tesh15, Oliver G. Pybus2,6,†, Marcio R. T. Nunes#1,15,‡, and Pedro F. C. Vasconcelos3,‡

**Affiliations**

1Center for Technological Innovation, Evandro Chagas Institute, Ministry of Health, Ananindeua, PA, 67030-000, Brazil
Acknowledgments

We thank Xavier de Lamballerie and John Lednicky for permission to include their unpublished ZIKV genomes in our analysis. We thank the Death Verification Service (SVO), Central Laboratories of Public Health (LACEN) and health departments of the Ceará State and Maranhão State, Brazil for collaboration. OGP is supported by the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 614725-PATHPHYLODYN. JL is supported by the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 268904-DIVERSITY. OGP received consulting fees from Metabiota Inc. between 2015-2016. This study is made possible in part by the generous support of the American people through the United States Agency for International Development (USAID) Emerging Pandemic Threats Program. The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government. SIH is funded by a Senior Research Fellowship from the Wellcome Trust (#095066), and grants from the Bill and Melinda Gates Foundation (OPP1119467, OPP1093011, OPP1106023, and OPP1132415). MRTN is funded as an associated Researcher in Public Health by the Evandro Chagas Institute, Brazilian Ministry of Health and as Researcher in Scientific productivity by CNPq (Brazilian National Council for Scientific and Technological Development) grant numbers 302032/2011-8, 200024/2015-9, and supported in part by the National Institute of Science and Technology for Viral Hemorrhagic Fevers. R.T. is funded by grant R24 AT 120942 from the U.S. National Institutes of Health. S.C.H. is supported by a Wellcome Trust grant (102427). T.A.B. and I.R. are supported by grants from the UK Medical Research Council (MR/L009528/1) and Wellcome Trust (090532/Z/09/Z). PFCV is supported by CNPq-National Agency for Scientific and Technologic Development (grants 573739/2008-0, 301641/2010-2, and 457664/2013-4). All samples were obtained from persons visiting local clinics or hospitalized by the Brazilian Ministry of Health personnel as part of dengue, chikungunya, and Zika fever surveillance activities. In these cases, patient consent is oral and not recorded. The study was authorized by the Coordination of the National Program for Dengue, Chikungunya, and Zika Control coordinated by Brazil’s Ministry of Health. The data are available at DRYAD: DOI:
The new ZIKV genomes reported in this study are deposited in GenBank under the accession numbers KU321639, KU365777 to KU365780, KU729217, and KU729218.

References and Notes


9. Pyke AT, Daly MT, Cameron JN, Moore PR, Taylor CT, Hewitson GR, Humphreys JL, Gair R. Imported Zika virus infection from the Cook Islands into Australia, 2014. PLOS Curr. 2014; doi: 10.1371/currents.outbreaks.4635a54dbffba2156f6bf276dc4965e


14. ECDC. Microcephaly in Brazil potentially linked to the Zika virus epidemic. ECDC. 2015

15. ECDC. Microcephaly in Brazil potentially linked to the Zika virus epidemic. ECDC. 2015


Science. Author manuscript; available in PMC 2016 October 15.


20. PAHO. Implications for public health in the Americas. Neurological syndrome, congenital malformations, and Zika virus infection. PAHO/WHO; 2015.


50. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WE, Wetter T, Suhai S. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in


Fig. 1. Time series and cartography of reported Zika virus and microcephaly cases in Brazil. (A) Number of suspected cases of ZIKV per week in 5596 municipalities in Brazil. The epidemic peaked from 12 to 18 July 2015 (n = 2791 cases). Letters indicate months. (B) Total incidence of ZIKV cases per 100,000 people in each federal state. Triangles indicate sampling locations of the sequences reported here; circles indicate locations of other genomes from Brazil [municipality of Natal in Rio Grande do Norte state (16) and an unknown municipality in Paraíba state (21)]. Red symbols indicate ZIKV genomes isolated from microcephaly cases. Federal states are indicated by 2-letter codes: PA: Para, MA: Maranhão, CE: Ceará, RN: Rio Grande do Norte, PB: Paraíba. Per capita incidences in each state were calculated using high-resolution gridded human population size datasets for Brazil (45). (C) Incidence of suspected microcephaly cases per 100,000 people in each federal state. Per capita incidences for each state were calculated as described for panel (B).
Fig. 2. Maximum likelihood phylogeny of ZIKV complete coding region sequences. Bootstrap scores are shown next to well-supported nodes and the phylogeny was mid-point rooted. A fully annotated tree is provided in Fig. S2. The American ZIKV outbreak clade is drawn as a narrow white triangle and is shown in detail in Fig. 3. Asterisks highlight the four internal branches that are ancestral to the American ZIKV lineage (see main text and Fig. S3). Correlation between the sampling date of each sequence and the genetic distance of that sequence from the root of a maximum likelihood phylogeny of the Asian genotype (correlation coefficient $R^2 = 0.997$). A molecular clock phylogeny of this data is shown in

*Science. Author manuscript; available in PMC 2016 October 15.*
Fig. 3. The Malaysian strain (HQ234499) sampled in 1966 is the oldest representative of the Asian genotype and falls on the regression line, indicating that it does not appear to be unusually divergent for its age. A similar analysis with the HQ234499 strain excluded is shown in fig. S5C.
Fig. 3. Timescale of the introduction of ZIKV to the Americas.

(A) Molecular clock phylogeny of the ZIKV outbreak lineage estimated from complete coding region sequences, plus 6 sequences (KJ634273, KU312315, KU312314, KU212313, KU646828, and KU646827) longer than 1500nt (available data as of 7th March 2016). For visual clarity, three basal sequences, HQ23499 (Malaysia, 1966), EU545988 (Micronesia, 2007) and JN860885 (Cambodia, 2010) are not displayed here (see Fig. S3). Gray horizontal bars represent 95% Bayesian credible intervals for divergence dates. A and B denote clades discussed in main text and numbers next to them denote posterior probabilities. Diamond
sizes represent, at each node, the posterior probability support of that node. Taxa are labeled with accession number, sampling location, and sampling date. Names of sequences generated in this study are underlined. (B) Posterior distributions of the estimated ages (TMRCAs) of clades A and B, estimated in BEAST software using the best-fitting evolutionary model (table S2). The time and duration of the three events (i-iii) discussed in the main text are shown. (C) Number of airline passengers from specific countries arriving in Brazil per month versus number of suspected cases of ZIKV in French Polynesia. The blue curve (left y axis) shows a polynomial fitting of the number of travelers (blue points) from countries with recorded ZIKV outbreaks between 2012 and 2015 (French Polynesia, Thailand, Indonesia, Malaysia, Cambodia, and New Caledonia) (supplementary materials section 6), aggregated across 20 Brazilian national airports. The purple bars represent weekly numbers of suspected ZIKV cases (right y axis) in French Polynesia (FP) from 30 October 2013 to 14 February 2014 (4).