

Epidemiology and evolution of Zika Virus in Minas Gerais, Southeast Brazil

Felipe C.M. Iani^{1,2,3}; Marta Giovanetti^{3,4}; Vagner Fonseca^{3,5,6}; William M. Souza^{2,7}; Talita E. R. Adelino^{1,3}; Joilson Xavier³; Jaqueline G. Jesus^{7,8}; Maira A. Pereira¹; Marcos V. F. Silva¹; Alana V. B. Costa¹; Erniria C. Silva¹⁰; Márcia C. O. Mendes¹⁰; Ana M.B. Filippis⁴; Carlos F. C. Albuquerque¹¹; André L. Abreu⁶; Marluce A. A. Oliveira^{1*}; Luiz C. J. Alcantara^{3,4*} & Nuno R. Faria^{2,8,12*}

Affiliations:

1. Laboratório Central de Saúde Pública, Fundação Ezequiel Dias, Belo Horizonte, Minas Gerais, Brazil; 2. Department of Zoology, University of Oxford, United Kingdom; 3. Laboratório de Genética Celular e Molecular, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; 4. Laboratório de Flavivírus, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil; 5. KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), College of Health Sciences, University of KwaZuluNatal, Durban 4001, South Africa; 6. Coordenação Geral dos Laboratórios de Saúde Pública/Secretaria de Vigilância em Saúde, Ministério da Saúde, Brasília, Distrito Federal, Brazil; 7. Centro de Pesquisa em Virologia, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brazil. 8. Instituto de Medicina Tropical, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil. 9. Departamento de Moléstias Infecciosas e Parasitárias, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil. 10. Secretaria de Estado de Saúde de Minas Gerais, Belo Horizonte, Brazil; 11. Organização Pan-Americana da Saúde/Organização Mundial da Saúde, Brasília-DF, Brazil; 12. MRC Centre for Global Infectious Disease Analysis, J-IDEA, Imperial College London, London, UK.

***Corresponding author:**

Nuno R. Faria (n.faria@imperial.ac.uk)

Luiz C. J. Alcantara (luiz.alcantara@ioc.fiocruz.br)

Marluce A. A. Oliveira (marluce.oliveira@funed.mg.gov.br)

1. Introduction

Zika virus (ZIKV) is a mosquito-borne virus with a 11kb positive-sense single-stranded RNA genome. ZIKV belongs to the *Flavivirus* genus (*Flaviviridae* family), the same virus family as dengue virus (DENV), and yellow fever virus (Simmonds et al., 2017). ZIKV was first isolated in 1947 from a sentinel monkey in the Zika forest, located in southeast Uganda, East Africa (Baud et al., 2017). ZIKV infection causes a self-limiting disease and up to 80% of cases are asymptomatic (Haby et al., 2018). Most of the symptomatic cases are characterized by an acute febrile illness with headache, myalgia, conjunctivitis, and/or maculopapular rash (Ioos et al., 2014;). However, ZIKV infection can cause severe neurological disease, such as microcephaly in newborns and Guillain-Barré syndrome in adults (Baud et al., 2017).

ZIKV is mainly transmitted to humans by the *Aedes aegypti* mosquitoes, the same vector of DENV, and Chikungunya virus (CHIKV) (Boyer et al 2018; Musso and Gubler, 2016). These viruses co-circulate in the same geographic area and cause clinically similar illnesses (Grubaugh et al., 2018). Therefore, the reliable diagnosis of ZIKV infection requires detecting ZIKV RNA through real-time reverse transcription-polymerase chain reaction (RT-qPCR). ZIKV can be classified into two geographically and phylogenetically distinct lineages, the African and Asian genotypes (Faye et al., 2013). The ZIKV Asian genotype has recently emerged as one of the most serious global public health threats (Baud et al., 2017); its first reported epidemic was described in 2007 in Micronesia, followed by outbreaks in 2013-2014 in several Pacific islands and 2015 in Africa (Musso and Gubler, 2016; Faye et al., 2020).

In April 2015, the first autochthonous ZIKV cases were confirmed in northeast Brazil (Ministério da Saúde, 2017). Until August 2020, ZIKV caused 294,713 cases in Brazil (Ministério da Saúde, 2018a, 2020) and 3,474 microcephaly cases associated with ZIKV infection until October 2019 (Ministério da Saúde, 2019). Previous studies investigated ZIKV's transmission in 12 out of 27 states of Brazil, including the federal district (Faria et al., 2016b, 2017; Giovanetti et al., 2020). Moreover, there is no information about the genetic diversity of ZIKV circulating in Minas Gerais, the second most populous Brazilian state, bordering Rio de Janeiro and Sao Paulo states, and located in the southeast region of the country. In this study, we combine epidemiological analysis and portable genome sequencing with data generated at *Fundação Ezequiel Dias* (FUNED), the Public Central Health Laboratory of Minas Gerais, to describe the ZIKV epidemic in Minas Gerais between 2015 to 2020.

2. Material and methods

2.1. Sample collection

Serum, urine, tissue samples from all patients with ZIKV symptoms attended by public health services in the Minas Gerais States were collected for molecular diagnostics and sent for testing at *Fundação Ezequiel Dias* (FUNED), the Public Central Health Laboratory of Minas Gerais. Sampled patients subjected to molecular diagnostics presented maculopapular rash and at least two symptoms, such as fever, polyarthralgia, peri-articular edema, and conjunctivitis (purulent or hyperemic). All

samples were processed under terms of Resolution 510/2016 of national ethical review board (*Comissão Nacional de Ética em Pesquisa*), under the auspices of the ZIBRA2 project (www.zibra2project.org/). The project was approved by the Pan American Health Organization Ethics Review Committee (PAHOERC) no. PAHO-2016-08-0029.

2.2. Nucleic acid isolation and RT-qPCR for ZIKV

Viral nucleic acid extraction was performed using the MagNA Pure 96 System (Roche Diagnostics, Switzerland) according to the manufacturer's instructions. Molecular diagnostic was performed by RT-qPCR against the prM target specific to ZIKV (using 5' FAM as the probe reporter dye) and GoTaq® 1-Step RT-qPCR System (Promega, USA), as previously described (Lanciotti et al., 2008). Cycle threshold (Ct) values were determined for all samples. All procedures were conducted in biological safety cabinets located in physically separated areas. Negative controls were used in all reactions.

2.3. ZIKV sequencing and consensus sequences

Genome sequencing was performed in samples with Ct values <35 and availability of epidemiological data, such as date of onset of symptoms, date of sample collection, sex, municipality of residence, and symptoms. A total of 10 ZIKV samples with PCR products yielding sufficient material (DNA concentration after clean-up being > 4ng/μL), were randomly selected for genome sequencing.

Positive samples were submitted to a cDNA synthesis protocol (Quick et al., 2017) using Superscript IV cDNA Synthesis Kit. Then, a multiplex tiling PCR was conducted with Q5 High Fidelity Hot-Start DNA Polymerase (New England Biolabs) using ZIKV sequencing primers scheme designed using Primal Scheme (<http://primal.zibra2project.org>) (Quick, Grubaugh et al. 2017). The thermocycling conditions involved 40 cycles; reaction conditions were previously reported in (Quick et al., 2017). Sequencing libraries were generated from the barcoded products using the Genomic DNA Sequencing Kit SQK-MAP007 and SQK-LSK208 and library quality was assessed by Qubit quantification after barcoding and adapter ligation steps. DNA was loaded onto R9.4 flow cells (Oxford Nanopore Technologies, United Kingdom). Raw files were basecalled using Guppy (Oxford Nanopore Technologies, United Kingdom), demultiplexed and trimmed using Porechop (<https://github.com/rrwick/Porechop>) and/or QCAT, and then mapped generating contigs (de novo) and aligned with the reference (GenBank accession number NC_035889.1) using Genome Detective (Vilsker et al., 2019).

2.4. Collation and sequence alignment of ZIKV-Asian complete genome datasets

Genotype of the newly generated genomes were identified using Genome Detective (Vilsker et al., 2019). New data was then appended to publicly available data for subsequent analysis. Two ZIKV

genome (coverage >50%) data sets were compiled: dataset 1 (n=474), comprised the data reported in this study (n=10) plus (n=464) publicly available ZIKV-Asian genotype genomes available until November 2019 in GenBank and after being filtered by genome coverage. The second one, dataset 2 (n=196), included only Brazilian strains (n=177 being 10 from this study) plus the oldest ZIKV Asian sequences (n=19). Sequence alignment was performed using MAFFT version 7 (Kato and Standley, 2013) and visually inspected in AliView version 1.26 (Larsson, 2014). Since recombination may impact evolutionary estimates before performing our phylogenetic reconstruction, we employed the 12 recombination detection methods available in RDP version 4 (Martin and Rybicki, 2000) to further search for evidence of recombination in our dataset. Moreover, no evidence of recombination was found.

2.5. Maximum likelihood analysis and clock signal estimation

Maximum likelihood (ML) trees were estimated using IQ-TREE 1.6.12 (Nguyen et al., 2014) under an GTR+F+R3 nucleotide substitution model for the dataset 1 and TIM2+F+R2 for the dataset 2, as indicated by ModelFinder implemented in IQ-TREE (Kalyaanamoorthy et al., 2017). Statistical robustness of tree topology was inspected using 1,000 bootstrap replicates; a bootstrap value >80% was considered strong statistical support. To estimate temporal signal in each dataset, sample collection dates were regressed against root-to-tip genetic distances obtained from the ML phylogenies using TempEst 1.5.3 (Rambaut et al., 2016). When precise sampling dates were not available, a precision of 1 month or 1 year in the collection dates was considered. For convenience Brazilian sequences have been grouped into states macro region for which those sequences were generated (Southeast: Minas Gerais, São Paulo, Rio de Janeiro; Northeast: Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Rio Grande do Norte; and North: Amazonas, Pará, Tocantins).

2.6. Dated phylogenetics

Time scaled phylogenetic trees were inferred using the BEAST version 1.10.4 statistical framework (Suchard et al., 2018). We used the codon-based SRD06 model (Shapiro et al., 2006) and a non-parametric Bayesian skyline coalescent model (Drummond and Rambaut, 2015) to model changes in effective population size over time and the uncorrelated relaxed molecular clock model (Drummond et al., 2006). Previous studies have demonstrated this combination to be the best fitting model combination for ZIKV in the Americas (Faria et al., 2017; Grubaugh et al., 2018, 2017). We computed 4 independent runs of 100 million MCMC steps, sampling parameters and trees every 10,000 steps. Convergence of MCMC chains was checked using Tracer v.1.7.1 (Rambaut et al., 2018). Maximum clade trees were summarized using TreeAnnotator version v1.10.4 after discarding 10% as burn-in. Phylogenetics analyses were performed using the Sagarana HPC cluster, CEPAD-ICB-UFMG.

2.7. Epidemiological data

We followed the case definition guidelines of the State Health Secretary of Minas Gerais (SES-MG). Specifically, a notified case was defined as a patient that presented maculopapular rash and at least two symptoms, such as fever, polyarthralgia, peri-articular edema, and conjunctivitis (purulent or hyperemic). A probable case was defined as a notified case, excluding the patients with laboratory diagnosis with negative results or diagnosed for other diseases. Confirmed cases were defined as patients with positive laboratory results for ZIKV or clinical-epidemiological criteria. Microcephaly incidence was estimated using the number of confirmed cases divided by the number of live births obtained by DATASUS (www.datasus.saude.gov.br). Incidences were calculated based on the estimated population of Minas Gerais State in 2019, as reported by the Brazilian Institute of Geography and Statistics (www.ibge.gov.br). Association between tested and confirmed cases, population size, incidence, and microcephaly cases in Minas Gerais state was determined by Spearman correlation analysis. Results were plotted using log10 transformed values after correlation analyses. All the statistics and maps were done using RStudio version 1.2.5033(RStudio Team, 2019).

2.8. Data availability

Epidemiological data and phylogenetic trees, XMLs are available on Zibra II Project website repository (<https://www.zibra2project.org/epidemiology-and-evolution-of-zika-virus-in-minas-gerais-southeast-brazil/>). ZIKV sequences from Minas Gerais State are available on GenBank (accession numbers: MT439638, MT439639, MT439640, MT439641, MT439642, MT439643, MT439644, MT439645, MT439646, MT439647).

3. Results

The first ZIKV RT-qPCR confirmed case in Minas Gerais state was reported on February 19, 2016 by the Ezequiel Dias Foundation (available on the *Gerenciador de Ambiente Laboratorial* - an electronic system maintained by the Ministry of Health), although if the official bulletin from the MG's Secretary of Health was only published in March 7, 2016 (Secretaria de Estado da Saúde de Minas Gerais, 2016). In an attempt to obtain evidence of ZIKV transmission before its first detection, we retrospectively assessed 513 samples collected from September 2015 to January 2016 from patients with acute febrile illness. Of these, we found 141 ZIKV RT-qPCR-positive samples (8.2% of all RT-qPCR confirmed cases) collected between December 6, 2015 and January 31, 2016, revealing that ZIKV was likely circulating unnoticed for more than 2 months in Minas Gerais before its first detection. (**Figure 1A**).

Time-series of the ZIKV RT-qPCR positive cases show two epidemic waves: the 'first ZIKV wave with 94% of all RT-qPCR confirmed cases (1,623 cases) between December 2015 to May 2016 with a peak of cases registered on March 2016 (with n=663, 41% of all RT-qPCR confirmed cases of the first ZIKV wave) (**Figure 1A**) and a second, smaller epidemic wave with 51 RT-qPCR confirmed cases from March to July 2017 (3% of all RT-qPCR confirmed cases detected during the study period) (**Figure 1B**).

We next evaluated ZIKV incidence across Minas Gerais regions. The *Norte de Minas* (north region) was the mesoregion with the highest incidence of RT-qPCR confirmed cases during the first ZIKV wave (**Figure 2 A**). The Ipatinga municipality accounted for the highest number of ZIKV cases confirmed by clinical-epidemiological criteria ($n=2,546$) and with RT-qPCR ($n = 79$), which reported an incidence of 29.99 per 100,000 inhabitants in 2016. Besides, the highest number of RT-qPCR confirmed cases was detected in Belo Horizonte city (state capital city) ($n = 486$) (**Supplementary Figure 1A**), that recorded an incidence of 19.34 per 100,000 inhabitants (**Supplementary Figure 1B**). Interestingly, the last ZIKV RT-qPCR confirmed cases in the Minas Gerais state were detected on July 31, 2017. ZIKV RT-qPCR confirmed cases were predominantly in adults (18 to 40 years old) (78.81%, $n = 1,358/1,723$). The median age was 30.28 years (interquartile range: 23-35), ranging from 1 day to 78 years. Samples were characterized by mean Ct of 29.3 (range 21.21 to 39.99) (**Figure 2B**). We found that municipality population size was moderately associated with ZIKV confirmed cases (Spearman's rank coefficient of 0.38, $p = 6e-08$). Moreover, we found a strong correlation between the number of total cases tested per municipality versus ZIKV confirmed (Spearman's rank coefficient of 0.8, $p < 2.2e-16$) (**Supplementary Figure 2A and B**).

Most RT-qPCR confirmed cases were in females (91.6%, $n=1,578$) (**Figure 2C**), most likely due to guidelines in testing. Noteworthy, we report one case in embryonic fragment of an abortion of an eight-week RT-qPCR positive pregnant woman (accession number GenBank: MT439640) and 43 RT-qPCR confirmed cases in children (0 to 12 years old). The ZIKV outbreak (2016-2017) in Minas Gerais caused 83 microcephaly cases confirmed by ZIKV infection in 43 municipalities, 16.9% (14/83) were reported in Belo Horizonte city (Ministério da Saúde, 2018b). The *MG Central mesoregion* had the highest 2016 microcephaly incidence 80.77 per 100,000 newborns (**Figure 2 D**). In the state, most microcephaly cases were reported in November 2016 ($n=22$), approximately eight months after the peak of the first ZIKV wave (**Figure 1a**). The Spearman correlation analysis showed an association between microcephaly and the number of newborns ($r=0.67$, $p = 1e-06$). A correlation was also observed between microcephaly and confirmed cases per municipality ($r=0.58$, $p = 0.0012$) (**Supplementary Figure 2 C and D**).

Using handheld nanopore technologies, we obtained 10 ZIKV genome sequences, (coverage range 60.5%-91.5%, mean=76.56%), five from Belo Horizonte, and other five from municipalities of Betim, Matozinhos, Montes Claros, Santa Luzia, and Teófilo Otoni. All samples sequenced in this study were from the first ZIKV wave in Minas Gerais state and they had mean Ct value of 29.28 (range 23.87-33.45). Mean threshold cycle value of the generated sequences, sequencing statistics and epidemiological data are detailed in Table 1.

To better understand the transmission dynamics of ZIKV in MG state we estimate an initial maximum likelihood (ML) phylogenetic tree (**Supplementary figure 3**) analysis on the 10 new sequences combined with another 464 publicly available ZIKV-Asian genotype genomes. Our estimated phylogeny identified that the newly ZIKV genomes obtained in this study fell within the ZIKV American clade (bootstrap score = 100) (**Supplementary figure 3**).

To assess the evolution of ZIKV in MG state, we performed Bayesian time-measured phylogenetic analysis using a molecular clock model on a dataset 2 (n=196) comprising the 10 new sequences from MG plus other Brazilian strains (n=167) plus the oldest ZIKV Asian sequences (n=19) sequences. A regression of genetic divergence from root to tip against sampling dates confirmed sufficient temporal signal ($r^2=0.734$) (**Supplementary figure 4**). Our maximum clade credibility (MCC) tree identified that at least three independent introductions of ZIKV occurred in MG state between end-October 2014 and early-April 2015 (see clades MG1 MG2 and MG3 in Figure 3B). We observed that MG2 clade includes isolate MT439646 which was also closely placed to isolates MT439644 and MT439643 in the ML tree with a bootstrap value of 0.91. MG2 internal nodes present low bootstrap values which might reflect sequence low quality and coverage. Furthermore, formation of clades MG1 and MG3 were also observed in the ML tree, which is consistent with our time-measured Bayesian phylogeny. We estimated the most recent common ancestor (TRMCA) for each clade with a 95% of high posterior density to be early-April 2015 (November 2014 to September 2015) for the MG1, end-December 2014 (August 2014 to May 2015) for the MG2, and end-October 2014 (June 2014 to February 2015) for the MG3. We also identified three isolates from 2016 outside those three main clades. Isolate MT439641, sampled in April 2016, clusters with sequences from Northeast Brazil. Isolate MT439647, sampled in April 2016, clusters with sequences from Northeast and Southeast Brazil, and isolate MT439645 sampled in March 2016 falls basally to a clade containing sequences from Southeast Brazil. Taken together these data suggest that multiple independent introduction events have occurred into MG state mainly from Southeastern (MG1 and MG2), and Northeastern (MG3) Brazilian states (**Figure 3 A, B**).

4. Discussion

In this study, combining genomics and epidemiological data we investigate the ZIKV outbreak in Minas Gerais State between 2015 - 2017. Epidemiological data reveal that two main waves were responsible for the ZIKV epidemic in MG state, the first and largest one registered in 2016, followed by a second smaller one in 2017, which peaked around March, during the rainy season, a period with climatic suitability for arbovirus transmissions, such as DENV in the state (Aguiar et al., 2014). Interestingly, the MG state as well as other bordering states such as São Paulo, Goiás and Espírito Santo, were not characterized by a third epidemic wave between 2018-2019 as previously reported for the Amazon region (Giovanetti et al., 2020; Secretaria de Estado da Saúde de Minas Gerais, 2020).

The generated genomic data allowed us to estimate the introduction date of the ZIKV-Asian lineage in MG to have occurred between end-October 2014 and early-April 2015, suggesting an undetected circulation of the virus for 16 months before the first reports of ZIKV transmission in the state (Secretaria de Estado da Saúde de Minas Gerais, 2016) following a pattern of cryptic transmission that have been already observed before for zika as well as for other mosquito-borne viruses such as dengue and chikungunya virus epidemics (Faria et al., 2017; Xavier et al., 2019; Faye et al., 2020). According to the Minas Gerais State Health Secretary epidemiological bulletin, ZIKV autochthony in

MG was reported on 8 of March (Secretaria de Estado da Saúde de Minas Gerais, 2016). Prior to that, only imported cases had been registered.

Moreover, our data suggest that the circulation of the ZIKV-Asian genotype in MG may resulted from at least three independent introduction events over 2014 and 2015 (Figure 3), which we infer to have occurred, during a period characterized by high climatic suitability for arbovirus transmission in the region (Aguilar et al., 2014). Further, we found evidence that two (MG1 and MG2) clades had originated from the southeast region whereas a third one (MG3 clade) appears to have originated directly from Northeast Brazil, that have played a significant role in the establishment and dissemination of ZIKV in the Americas (Faria et al., 2016a)

Based on retrospective investigations reported by Ministry of Health, it was confirmed that ZIKV was the etiological agent of six microcephaly cases in December 2015 (Ministério da Saúde, 2018b), two months before the first ZIKV cases reported in Minas Gerais state on February 2016 by FUNED. These findings also corroborate with our estimates reinforcing the idea of a likely ZIKV cryptic transmission before its first detection in the state.

Together, our results shed light on the epidemiological dynamics of the ZIKV-Asian genotype into MG state, showing that genomic data generated by portable sequencing technology can be employed to assist public health laboratories in monitoring and understanding the diversity of circulating mosquito-borne viruses.

Declaration of Competing Interest

This work does not have any relationships with business related issues, and no conflict of interest exists in the submission of this manuscript.

Acknowledgements

We acknowledge The Brazilian Ministry of Health for the microcephaly cases data disponibilization in special *Wanderson Oliveira* and *Sulamita Brandão Barbiratto* and *Darlan Candido* for the initial preparation of sequence dataset. We also thanks Minas Gerais State Scretariat for Health for providing the epidemiological data and the Ezequiel Dias Foundation for the facilities availability.

Funding: This work was supported by Decit/SCTIE/BrMoH/CNPq (440685/2016-8; 421598/2018-2), by CAPES(88887.130716/2016-00) and by the European Union's Horizon 2020 Research and Innovation Programme under ZIKAlliance Grant Agreement no. 734548. F.C.M.I and J.X. are supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. F.C.M.I is also supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). M.G. is supported by the Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ. N.R.F. is supported by a Wellcome Trust and Royal Society Sir Henry Dale Fellowship (204311/Z/16/Z). N.R.F is also supported by a Medical Research Council-São Paulo Research Foundation (FAPESP) CADDE partnership award (MR/S0195/1 and FAPESP 18/14389-0) (<http://caddecentre.org/>). V.F. is supported by the South African Medical Research Council (MRC- RFA-UFSP-01-2013/UKZN HIVEPI) and the NIH H3Abi-oNet network, which is an initiative of the Human Health and Heredity in Africa Consortium (H3Africa). W.M.S. is supported by the São Paulo Research Foundation, Brazil (2017/13981-0 and 2019/24251-9). J.G.J. s supported by FAPESP (2018/17176-8 and 2019/12000-1, 18/14389-0).

References

- Aguiar M, Rocha F, Pessanha JEM, Mateus L, Stollenwerk N. Carnival or football, is there a real risk for acquiring dengue fever in Brazil during holidays seasons? *Sci Rep* 2014;5:1–12. <https://doi.org/10.1038/srep08462>.
- Baud D, Gubler DJ, Schaub B, Lanteri MC, Musso D. An update on Zika virus infection. *Lancet* 2017;390:2099–109. [https://doi.org/10.1016/S0140-6736\(17\)31450-2](https://doi.org/10.1016/S0140-6736(17)31450-2).
- Boyer S, Calvez E, Chouin-Carneiro T, Diallo D, Failloux A. An overview of mosquito vectors of Zika virus. *Microbes and Infection* 2018;20:646–60. <https://doi.org/10.1016/j.micinf.2018.01.006>.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* 2006;4:699–710. <https://doi.org/10.1371/journal.pbio.0040088>.
- Drummond AJ, Rambaut A. Bayesian evolutionary analysis by sampling trees. *Bayesian Evol Anal with BEAST* 2015;8:79–96. <https://doi.org/10.1017/CBO9781139095112.007>.
- Faria NR, Quick J, Claro IM, Thézé J, De Jesus JG, Giovanetti M, et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* 2017;546:406–10. <https://doi.org/10.1038/nature22401>.
- Faria NR, Sabino EC, Nunes MRT, Alcantara LCJ, Loman NJ, Pybus OG. Mobile real-time surveillance of Zika virus in Brazil. *Genome Med* 2016a;8:97. <https://doi.org/10.1186/s13073-016-0356-2>.
- Faria NR, Do Socorro Da Silva Azevedo R, Kraemer MUG, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: Early epidemiological and genetic findings. *Science* (80-) 2016b;352:345–9. <https://doi.org/10.1126/science.aaf5036>.
- Faye Oumar, Faye Ousmane, Diallo D, Diallo M, Weidmann M, Sall AA. Quantitative real-time PCR detection of Zika virus and evaluation with field-caught Mosquitoes. *Virol J* 2013;10:1. <https://doi.org/10.1186/1743-422X-10-311>.
- Faye O, de Lourdes Monteiro M, Vrancken B, Prot M, Lequime S, Diarra M, et al. Genomic Epidemiology of 2015–2016 Zika Virus Outbreak in Cape Verde. *Emerg Infect Dis.* 2020;26(6):1084–1090. <https://dx.doi.org/10.3201/eid2606.190928>
- Giovanetti M, Faria NR, Lourenço J, Goes de Jesus J, Xavier J, Claro IM, et al. Genomic and Epidemiological Surveillance of Zika Virus in the Amazon Region. *Cell Rep* 2020;30:2275–2283.e7. <https://doi.org/10.1016/j.celrep.2020.01.085>.
- Grubaugh ND, Faria NR, Andersen KG, Pybus OG. Genomic Insights into Zika Virus Emergence and Spread. *Cell* 2018;172:1160–2. <https://doi.org/10.1016/j.cell.2018.02.027>.

Grubaugh ND, Ladner JT, Kraemer MUG, Dudas G, Tan AL, Gangavarapu K, et al. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* 2017;546:401–5. <https://doi.org/10.1038/nature22400>.

Haby MM, Pinart M, Elias V, Reveiz L. Prevalence of asymptomatic Zika virus infection: A systematic review. *Bull World Health Organ* 2018;96:402-413D. <https://doi.org/10.2471/BLT.17.201541>.

loos S, Mallet HP, Leparç Goffart I, Gauthier V, Cardoso T, Herida M. Current Zika virus epidemiology and recent epidemics. *Med Mal Infect* 2014;44:302–7. <https://doi.org/10.1016/j.medmal.2014.04.008>.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–9. <https://doi.org/10.1038/nmeth.4285>.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 2013;30:772–80. <https://doi.org/10.1093/molbev/mst010>.

Lanciotti RS, Kosoy OL, Laven JJ, Velez JO, Lambert AJ, Johnson AJ, et al. Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. *Emerg Infect Dis* 2008;14:1232–9. <https://doi.org/10.3201/eid1408.080287>.

Larsson A. AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 2014;30:3276–8. <https://doi.org/10.1093/bioinformatics/btu531>.

Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics*. 2000 Jun;16(6):562-3. doi: 10.1093/bioinformatics/16.6.562. PMID: 10980155.

Ministério da Saúde. Monitoramento dos casos de arboviroses urbanas transmitidas pelo Aedes. *Bol Epidemiológico Arboviroses* 2020;51:1–13.

Ministério da Saúde. Síndrome congênita associada à infecção pelo vírus Zika. vol. 50 (n. esp. 2019).

Ministério da Saúde. Boletim Epidemiológico 47: Situação epidemiológica da infecção pelo vírus Zika no Brasil, de 2015 a 2017. *Bol Epidemiológico S* 2018a;49:1–12.

Ministério da Saúde. Registro de Eventos em Saúde Pública (RESP-Microcefalia) 2018b. <http://www.resp.saude.gov.br/microcefalia#/painel> (accessed November 6, 2018).

Ministério da Saúde. Boletim Epidemiológico Secretaria de Vigilância em Saúde – Ministério da Saúde - até a Semana Epidemiológica 52, 2016. vol. 48. 2017.

Musso D, Gubler DJ. Zika virus. *Clin Microbiol Rev* 2016. <https://doi.org/10.1128/CMR.00072-15>.

Nguyen L, Schmidt HA, Haeseler A Von, Minh BQ. IQ-TREE : A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies 2014;32:268–74. <https://doi.org/10.1093/molbev/msu300>.

Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 2017;12:1261–6. <https://doi.org/10.1038/nprot.2017.066>.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol* 2018;67:901–4. <https://doi.org/10.1093/sysbio/syy032>.

Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2016;2:vew007. <https://doi.org/10.1093/ve/vew007>.

RStudio Team. RStudio: Integrated Development Environment for R 2019.

Secretaria de Estado da Saúde de Minas Gerais. Boletim Epidemiológico. 2020.

Secretaria de Estado da Saúde de Minas Gerais. Boletim epidemiológico de monitoramento dos casos de Dengue, Febre Chikungunya e Febre Zika. Nº 10, Semana Epidemiológica 11, 15/03/2016. 2016.

Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol* 2006;23:7–9. <https://doi.org/10.1093/molbev/msj021>.

Simmonds P, Becher P, Bukh J, Gould EA, Meyers G, Monath T, et al. ICTV virus taxonomy profile: Flaviviridae. *J Gen Virol* 2017;98:2–3. <https://doi.org/10.1099/jgv.0.000672>.

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 2018;4:1–5. <https://doi.org/10.1093/ve/vey016>.

Vilsker M, Moosa Y, Nooij S, Fonseca V, Ghysens Y, Dumon K, et al. Genome Detective: An automated system for virus identification from high-throughput sequencing data. *Bioinformatics* 2019;35:871–3. <https://doi.org/10.1093/bioinformatics/bty695>.

Xavier J, Giovanetti M, Fonseca V, Thézé J, Gräf T, Fabri A, et al. Circulation of chikungunya virus East/Central/ South African lineage in Rio de Janeiro, Brazil. *PLoS One* 2019;14:1–14. <https://doi.org/10.1371/journal.pone.0217871>.

Table 1: Sequencing and epidemiological statistics.

Accession number	Lab_id	Coverage (%)	Reads	Depth of Coverage	NT (Identity (%))	AA (Identity (%))	I/D/M/F	Ct value	Sample Type	Municipality	Collection Date	Onset date	Age	Sex
MT439639	2032/16 M	91.5	68415	4130.4	99.1	98.7	0/12	33.45	Serum	Teofilo Otoni	11/04/16	11/04/16	28	Female
MT439643	BH 06	87.5	4285	217.7	99.5	99.5	0/5	26.96	Serum	Belo Horizonte	12/04/16	NA	28	Female
MT439642	BH 05	84.2	5035	268.4	99.5	99.5	0/5	26.05	Serum	Belo Horizonte	06/04/16	05/04/16	17	Female
MT439638	1295/16 M	80.4	54258	3826.4	98.9	98.3	0/15	33.14	Serum	Belo Horizonte	24/03/16	21/03/2016	27	Female
MT439644	BH 07	79.8	3524	214.2	99.5	99.2	0/3	27.5	Serum	Santa Luzia	20/04/16	18/04/16	25	Female
MT439641	BH 02	73.4	2053	126.1	99.5	99.5	0/1	23.87	Serum	Matozinhos	05/04/16	01/04/16	37	Female
MT439647	BH 21	72.1	4383	391.7	99.5	99.2	0/7	30.04	Serum	Belo Horizonte	14/04/16	13/03/16	18	Female
MT439640	1377/16 M	69.1	21624	2277.7	98.4	97.8	15/4	32.09	Tissue	Belo Horizonte	24/03/16	Abortion	NA	NA
MT439645	BH 16	67.1	2745	208.5	99.6	99.4	0/3	29.8	Serum	Betim	18/03/16	16/03/16	30	Female
MT439646	BH 19	60.5	4173	455.1	99.4	98.8	1/2	29.95	Serum	Montes Claros	06/04/16	04/04/16	23	Female

Figures captions:

Figure 1. Temporal distribution of Zika positive and microcephaly cases in Minas Gerais over 2016-2018. (A) ZIKV RT-qPCR confirmed cases (bars) and microcephaly ZIKV associated cases (red points) per month in Minas Gerais state, southeast Brazil. (B) Comparative bar chart of the number of ZIKV positive and negative RT-qPCR cases per month in Minas Gerais state.

Figure 2: (A) Minas Gerais mesoregion map showing 2016 Zika virus incidence per 100,000 inhabitants and the origin location of novel sequences; (B) RT-qPCR confirmed cases by age groups and RT-qPCR cycle threshold; (C) RT-qPCR test results by sex and age; (D) 2016 microcephaly associated with Zika virus infection incidence per live birth Minas Gerais mesoregion map and the origin location of novel sequences.

Figure 3: Phylogenetic analysis of the introduction of Zika virus to Minas Gerais state. A) Dynamics of ZIKV introduction events in Minas Gerais state. Dates of introduction events were estimated from sequence data using a phylogenetic approach. B) Maximum clade credibility phylogeny estimated from Zika virus genomes with a molecular clock phylogenetic approach. Sequences are coloured according to sampling location. Clade A contains the MG Zika virus (red) and other closely related sequences from Brazil (green, blue and yellow indicate sampling location in Brazil). Clade posterior probabilities are shown at well-supported nodes.

Supplementary files

Supplementary Figure 1: A) Map of ZIKV RT-qPCR positive cases tested in Minas Gerais in 2016 per municipality; B) Map of Minas Gerais state showed ZIKV incidence per municipality per 100,000 inhabitants

Supplementary Figure 2: Spearman correlation of (A) the municipality population by Zika virus RT-qPCR confirmed cases with a coefficient of 0.38 and $p=6e-08$; (B) total cases tested per municipality versus positive cases with a coefficient of 0.8 and $p=2.2e-16$; (C) correlation between microcephaly associated with Zika virus infection and the number of newborns with a coefficient of 0.67 and $p=1e-06$; (D) microcephaly associated with Zika virus infection and RT-qPCR confirmed cases per municipality with a coefficient of 0.58 and $p=0,0012$. Y-axis values (A), (B), (C) and (D) were log 10 transformed after correlation analysis for easy chart reading.

Supplementary Figure 3: Maximum likelihood phylogeny was estimated with 474 ZIKV-Asian genotype genome sequences. Colors represent different locations. Scale bar represents expected substitutions per nucleotide site.

Supplementary Figure 4: Genetic divergence regressed against date of sample collection for ZIKV strains. Colors represent different locations.