First complete genome sequences of Zika virus from Midwest Brazil, 2015.

Carla Julia da Silva Pessoa Vieira, Laís Ceschini Machado, Lindomar José Pena, Roberta Vieira de Moraes Bronzoni, Gabriel da Luz Wallau.

Author affiliations: Health Sciences Institute, Federal University of Mato Grosso, Sinop, Mato Grosso, Brazil (C.J.S.P. Vieira, R.V.M. Bronzoni); Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Recife, Pernambuco, Brazil (L.C. Machado, L.J. Pena, G.L. Wallau)

Address for correspondence: Gabriel da Luz Wallau, Entomology Department, Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Av. Professor Moraes Rego, s/n, Cidade Universitária, 50.740-465 Recife, Pernambuco Brazil; email: gabriel.wallau@cpqam.fiocruz.br; gabriel.wallau@gmail.com

Abstract: Zika virus (ZIKV) has been intensively studied in Brazil since 2015-2016 epidemics, but little is known about the virus in Midwest region of the country. We report here two ZIKV complete genomes, which were isolated during arboviral surveillance in Sinop city, southern border of the Amazonian forest, Midwest Brazil, 2015.

Zika virus (ZIKV) is a mosquito-borne arbovirus that has a positive-sense single-stranded RNA genome. This virus belongs to the *Flaviviridae* family, *Flavivirus* genus, and has recently emerged as one of the most serious global public health threat. Symptoms are usually mild and up to 80% of ZIKV infections can be asymptomatic (1). However, ZIKV has been associated to neurological complications such as congenital Zika syndrome in newborns, Guillain-Barré syndrome and other neurological complications in adults (2, 3, 4).

ZIKV has been intensively studied since the 2015 epidemics in South America following the virus global spread (5, 6, 7). Although the study of full-length viral genomic sequences of clinical isolates can lead to a better understanding of molecular epidemiology, phylogeny and viral virulence, currently there are few available ZIKV complete sequences especially from regions of difficult access in Brazil. We report here a genomic investigation and phylogenetic study based on complete ZIKV genomes from patients living in Sinop, a city located 503 km north of the capital Cuiabá, in the State of Mato Grosso, Midwest region of Brazil.

We investigated 63 serum samples from patients presenting clinical symptoms of dengue-like infection, during an arboviral surveillance, from December 2015 to February 2016, in Sinop. These samples were tested for flaviviruses and alphaviruses by Multiplex-Nested RT-PCR (online Technical Appendix). Samples collection and research use was approved by the Ethics Committee from Júlio Müller University Hospital – UFMT (288.172/2013).

A total of six samples were positive for ZIKV by Nested-PCR. Two PCR-positive samples (BR/Sinop/H355/2015 and BR/Sinop/H366/2015) were amplified by real-time PCR, using primers, probe and PCR conditions as described (8) and sequenced on the MiSeq (Illumina) platform at the Technological Platform Core (online Technical Appendix). One sample (BR/Sinop/H355/2015) belonged to a 41-year-old female patient presenting petechial rash, back pain and arthralgia, and normal leukocyte ($4,530/\text{mm}^3$) and platelet count ($314,000/\text{mm}^3$). The other sample (BR/Sinop/H366/2015) was from a 30-year-old pregnant patient presenting fever, petechial rash, itch, diffuse myalgia and arthralgia, and normal leukocyte ($6,390/\text{mm}^3$) and platelet count ($140,000/\text{mm}^3$).

The entire ZIKV genomes from these samples were obtained with average coverage depth ranging from 280 to 9282 (Technical Appendix Table). Phylogenetic reconstruction of the obtained genomes showed that strain isolated from Sinop patients clustered within the Asian lineage (Figure, panel A) along with several other draft ZIKV genomes from the 2015-2016 epidemics. More specifically, one complete genome sequenced here clustered in a clade along with several other genomes from the

North and Southeast Brazilian regions and genomes from Central America (355 sample - Figure, panel B, clade A) and the 366 genome clustered within another clade along with samples from Northeast and Southeast Brazilian regions as well as genomes from Central America (Figure, panel B, clade B). In addition, we could observe that the most closely related ZIKV genomes were detected in patients from the states of Tocantins, east border state of Mato Grosso, and from Paraíba state, located in the Northeast region, an area highly affected by the ZIKV epidemics (Figure, panel C).

We also characterized a number of nonsynonymous nucleotide mutations in the genomes obtained. We detected the previously reported amino acid mutations in one structural and four of the nonstructural proteins, including the A983V mutation NS1, which has been implicated in immune evasion and is conserved in all Asian strains detected after the French Polynesia outbreak (9) and a pre-M mutation (S139N) which has been associated with microcephaly in mice (10). In addition, we also found mutations specific for our samples: a NS2B mutation (I1398V), a NS3 mutation (T2068M) and a NS4B mutation (S2456L) (Technical Appendix Table).

The largest incidence of Zika in the Mato Grosso state occurred in 2015 (11). From 2015 to 2018, 6920 cases of DENV, 1392 of ZIKV, 12 of CHIKV were reported in the city (12). This report supports that the ZIKV circulating in this remote region of Brazil belongs to the Asian genotype and more specifically to two different clades found in other Brazilian states. Moreover, the close clustering with ZIKV sequences available from other Brazilian states which are far apart from Mato Grosso state highlights the rapid expansion of these strain between distant Brazilian regions. Lastly, we found two of ZIKV amino acid mutations that were associated with the virus pathogenicity and conserved in all ZIKV genomes obtained after 2014 and several other amino acid changes which their role remains unknown. Such findings highlight the need of rapid and efficient deployment of molecular surveillance in affected areas of difficult access (5).

References

1. Haby MM, Pinart M, Elias V, Reveiz L. Prevalence of asymptomatic Zika virus infection: a systematic review. Bull World Health Organ. 2018;96:402–13.
<http://dx.doi.org/10.2471/BLT.17.201541>
2. Depoux A, Philibert A, Rabier S, Philippe HJ, Fontanet A, Flahault, A. A multi-faceted pandemic: a review of the state of knowledge on the Zika virus. Public Health Rev. 2018;39:10.
<http://dx.doi.org/10.1186/s40985-018-0087-6>
3. Miner JJ, Diamond MS. Zika Virus Pathogenesis and Tissue Tropism. Cell Host & Microbe, 2017;21(2), 134–142.
<https://doi.org/10.1016/j.chom.2017.01.004>
4. Krauer F, Riesen M, Reveiz L, Oladapo O T, Martínez-Vega R, Porgo TV, ... Low, N. Zika Virus Infection as a Cause of Congenital Brain Abnormalities and Guillain–Barré Syndrome: Systematic Review. PLOS Medicine, 2017;14(1), e1002203.
<https://doi.org/10.1371/journal.pmed.1002203>
5. Lowe R, Barcellos C, Brasil P, Cruz O, Honório N, Kuper H, et al. The Zika Virus Epidemic in Brazil: From Discovery to Future Implications. Int J Environ Res Public Health. 2018;15:96–113.
<http://dx.doi.org/10.3390/ijerph15010096>
6. Faria NR, Quick J, Claro IM, Théze J, de Jesus JG, Giovanetti M, ... Pybus OG. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature, 2017;546(7658), 406–410.
<https://doi.org/10.1038/nature22401>

7. Campos TDL, Durães-Carvalho R, Rezende AM, de Carvalho OV, Kohl A, Wallau GL, Pena LJ. Revisiting Key Entry Routes of Human Epidemic Arboviruses into the Mainland Americas through Large-Scale Phylogenomics. *International Journal of Genomics*, 2018, 1–9.

<https://doi.org/10.1155/2018/6941735>

8. 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://dx.doi.org/10.3201/eid1408.080287>

9. Liu Y, Liu J, Du S, Shan C, Nie K, Zhang R, et al. Evolutionary enhancement of Zika virus infectivity in *Aedes aegypti* mosquitoes. *Nature*. 2017;545(7655):482–6.

<http://dx.doi.org/10.1038/nature22365>

10. Yuan L, Huang X-Y, Liu Z-Y, Zhang F, Zhu X-L, Yu J-Y, et al. A single mutation in the prM protein of Zika virus contributes to fetal microcephaly. *Science*. 2017;358(6365):933–6.

<http://dx.doi.org/10.1126/science.aam7120>

11. Ministério da Saúde. Boletins Epidemiológicos – Secretaria de Vigilância em Saúde. Monitoramento dos casos de dengue, febre de chikungunya e febre pelo vírus Zika. 2018;47(33):10 [cited 2018 Jun 25]

<http://portalms.saude.gov.br/boletins-epidemiologicos>

12. Secretaria de Estado de Saúde de Mato Grosso. Boletins Epidemiológicos. Dengue, febre de chikungunya e febre pelo vírus Zika. 2018;1(6):6 [cited 2018 Jun 25]

<http://www.saude.mt.gov.br/dengue/arquivos/526/documentos>

Legend Figure - Phylogenetic reconstruction and positioning of ZIKV genomes obtained in this study using maximum likelihood. A - Full phylogeny including all available ZIKV complete and draft genomes > 9kb. B - Phylogenetic analysis of a subset of ZIKV sequences from the full ZIKV tree focusing on clades where the two ZIKV genomes obtained in this study clustered. C - Distribution of the number of ZIKV genomes sequenced per state and colored edges showing the possible route of ZIKV spread reaching Sinop based on the most closely related ZIKV genome from the previous phylogenetic analysis. Tip colors in panel B follow state colors in panel C. aLRT SH-like branch support of key clades are depicted with arrows in panels A and B.

First complete genome sequences of Zika virus isolated in the Brazilian Amazon Basin, 2015.

Technical Appendix

Laboratory testing

Total RNA was extracted from 140 µL of patients' sera using QIAmp viral RNA kit, followed by a reverse transcription using random primer, and Multiplex-Nested- RT-PCR using specific primers targeting NS5 and nsP1 regions, respectively (1). Briefly, we designed a new primer for ZIKV (CACTGGCCTCCTAGGCCCGTCCAT) and add it to the Multiplex-Nested- PCR for flaviviruses, along with DENV 1-4. We add CHIKV primer (TAGAGCAGGAAATTGATCCC) previously designed (2) to the Multiplex-Nested- PCR for alphaviruses.

Genome sequencing

We used primers designed by the ZIBRA project (3) following as described (4) on the cDNA generated from the total RNA extracted directly from each sample. Samples were sequenced on the MiSeq (Illumina) platform at the Technological Platform Core at the Aggeu Magalhães Institute (IAM). Trimming and quality filtering of the recovered reads were performed with Trimmomatic v 0.36 (5) and quality checked with FastQC (<http://www.bioinformatics.babra-ham.ac.uk/projects/fastqc/>). These were subsequently mapped against the ZIKV BRPE243/2015 reference genome (KX197192) (6) using Bowtie 2 (7). The obtained sequences of the ZIKV genome have been deposited in GenBank with the accession number: MH513598 and MH513599.

Phylogenetic Analyses

Multiple sequence alignment was performed by using MAFFT version 7 (<http://mafft.cbrc.jp/alignment/software/>); maximum-likelihood (ML) was determined by using PhyML version 3.0 (<http://www.atgc-montpellier.fr/phyml/>). Coding regions corresponding to the complete genomes from Sinop were aligned with all published and available near-complete Zika virus genomes (>9,000 nucleotides), belonging to the African and Asian genotype, totalizing 408 genomes, collected on GenBank. The ML phylogeny was reconstructed by using the best-fit general time-reversible (GTR) model with 4 gamma substitution rate categories (+ G) distributed with invariant sites (+ I) (GTR+G+I); for the tree search operation, we used SPR and statistical support for phylogenetic nodes was assessed by using aLRT SH-like. We also used FigTree v1.4.2 (8) for visualization and figure generation.

References

1. Bronzoni RVM, Baleotti FG, Maria R, Nogueira R, Nunes M, Figueiredo, L.T. Duplex Reverse Transcription-PCR Followed by Nested PCR Assays for Detection and Identification of Brazilian Alphaviruses and Flaviviruses. *J Clin Microbiol.* 2005;43(2), 696–702.
<http://dx.doi.org/10.1128/JCM.43.2.696-702.2005>.
2. Hasebe F, Parquet MC, Pandey BD, Mathenge EGM, Morita K, Balasubramaniam V, et al. Combined detection and genotyping of Chikungunya virus by a specific reverse transcription-polymerase chain reaction. *J Med Virol.* 2002;67(3), 370–4.
<https://dx.doi.org/10.1002/jmv.10085>.
3. Faria NR, Sabino EC, Nunes MRT, Alcantara LCJ, Loman NJ, Pybus OG. Mobile real-time surveillance of Zika virus in Brazil. *Genome Med.* 2016;8:97.
<https://dx.doi.org/10.1186/s13073-016-0356-2>.

4. Guedes DR, Paiva MH, Donato MM, Barbosa PP, Krokovsky L, Rocha SWS. Zika virus replication in the mosquito *Culex quinquefasciatus* in Brazil. *Emerg Microbes Infect.* 2017;6:e69. <https://dx.doi.org/10.1038/emi.2017.59>
5. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014; 30:2114–2120. <https://dx.doi.org/10.1093/bioinformatics/btu170>
6. Donald CL, Brennan B, Cumberworth SL, Rezelj VV, Clark JJ, Cordeiro MT, et al. Full genome sequence and sfRNA interferon antagonist activity of Zika VIRUS from Recife, Brazil. *PLoS Negl Trop Dis.* 2016;10:e0005048. <https://dx.doi.org/10.1371/journal.pntd.0005048>
7. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357–359. <https://dx.doi.org/10.1038/nmeth.1923>
8. Rambaut A. FigTree: tree figure drawing tool version 1.4.3. Edinburgh (SCT): Institute of Evolutionary Biology, University of Edinburgh. 2014.

Technical Appendix Table: Number of reads, quality and coverage for the genomes reported in this study. Genome coverage (%) was calculated relative to sequence KX197192 (10807 nt long).

Isolate ID	Genbank accession code	No. total reads	No. mapped reads (%)	Depth of coverage	Genome length (nt)	Amino acid mutations and protein location*
H355	MH513598	676,945	40,478 (5.98%)	280.92	10640	M1143V NS1 I1398V NS2B T2068M NS3 S2456L NS4B
H366	MH513599	1,454,943	1,337,597 (91.93%)	9,282.85	10658	M1143V NS1

Depth of coverage = Number of mapped reads x 75 (read length) / reference genome length (10,807).

*Positions considering the reference genome polypeptide

A.

B.

C.