Discovery of a persistent Zika virus lineage in Bahia, Brazil

Samia N. Naccache*[1,2], Julien Thézé*[3], Silvia I. Sardi[4], Sneha Somasekar[1,2], Alexander L. Greninger[5], Antonio C. Bandeira[6], Gubio S. Campos[4], Laura B. Tauro[7], Nuno R. Faria[3,8], Oliver G. Pybus[3], and Charles Y. Chiu[1,2,9**]

Affiliations:

[1]Department of Laboratory Medicine, University of California, San Francisco, CA 94107, USA.

[2]UCSF-Abbott Viral Diagnostics and Discovery Center, San Francisco, CA 91407, USA.

[3]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

[4]Laboratory of Virology, Health Institute of Science, Federal University of Bahia, Salvador, Bahia, Brazil

[5]Department of Laboratory Medicine, University of Washington, Seattle, CA 98195, USA.

[6]Alianca Hospital, Salvador, Bahia, Brazil

[7]Center of Research Gonzalo Moniz, Foundation Fiocruz, Salvador, Bahia, Brazil

[8]Center for Technological Innovation, Evandro Chagas Institute, Ministry of Health, Ananindeua, PA 67030-000.

[9]Department of Medicine, Division of Infectious Diseases, University of California, San Francisco, CA 94107

*Co-first authors

**Corresponding author: Charles Chiu, MD/PhD, charles.chiu@ucsf.edu

Keywords: Zika virus, ZIKV, flaviviruses, Bahia, Brazil, mosquito-borne infections, outbreak surveillance, metagenomic next-generation sequencing (mNGS), viral genome assembly, capture probe enrichment, phylogenetic analysis, molecular clock

Running Title: Persistent ZIKV lineage in Bahia, Brazil

Abstract

Metagenomic next-generation sequencing coupled with capture probe enrichment was used to recover 11 whole and partial Zika virus (ZIKV) genomes from patients in Bahia, Brazil from April 2015 to January 2016, where the majority of suspected Brazilian ZIKV cases have been reported. Phylogenetic reconstructions and molecular clock analyses using the newly generated data uncovered the existence of a Bahia-specific ZIKV lineage sharing a common ancestor in mid-2014, indicating sustained circulation of this strain in Bahia since that date.

Introduction

Zika virus is an arthropod-borne RNA virus primarily transmitted by mosquitoes of the *Aedes* species, although transmission through blood transfusion and sexual contact has also been described (1). ZIKV can be classified in two genotypes: the African genotype, found only in the African continent, and the Asian genotype, associated with outbreaks in Southeast Asia, several Pacific islands and, more recently, in the Americas (2). In May of 2015, Brazil reported the first detection of autochthonous cases of ZIKV in the northeast of the country (3, 4). Since then, transmission has been confirmed in 22 of the 27 federal states (5).

The rapid geographic expansion of ZIKV transmission, together with its proposed association with microcephaly and congenital abnormalities (6), demand a rapid increase in

molecular surveillance in the most affected areas. This is particularly relevant for regions where other mosquito-borne viruses, particularly dengue virus (DENV) and chikungunya virus (CHIKV), are known to co-circulate with ZIKV (2), as surveillance based on clinical symptoms alone can lead to misdiagnosed cases. Accurate genetic characterization of circulating strains can also facilitate determination of the origin and potential spread of ZIKV infection in travelers returning from endemic countries. Previous analyses have suggested that ZIKV was introduced in the Americas at least one year before its initial detection in Brazil (1). Bahia state comprised 93% of suspected ZIKV cases in Brazil in 2015 (2), and has reported cases of ZIKV-associated fetal microcephaly, although no genetic information aside from one complete genome is available from the region to date (2, 7). Here we report epidemiological findings obtained using 11 new complete and partial ZIKV genomes from 15 clinical cases seen at Hospital Aliança in Salvador, Bahia, between April 2015 and January 2016.

Materials and Methods

Symptomatic patients were diagnosed with acute ZIKV infection by positive qualitative RT-PCR testing from serum using primers targeting the NS5 gene (8). Metagenomic next-generation sequencing (NGS) libraries were constructed following DNase treatment of nucleic acid extracts as previously described (9, 10), followed by pathogen detection using the SURPI bioinformatics pipeline (11). Clinical samples with titers exceeding 10^4 copies per mL by ZIKV quantitative RT-PCR using the QuantiTect SYBR Green RT-PCR kit (Qiagen) and primers targeting the fiber gene (12) produced sufficient metagenomic data to permit complete ZIKV genome assembly. For the remaining samples at lower titers, metagenomic NGS libraries were enriched for ZIKV sequences using a set of 299 XGen biotinylated lockdown capture probes (IDT Technologies) designed to tile across all ZIKV genomes (>10 kB) in the National Center for Biotechnology

Information (NCBI) GenBank database as of February 2016, and curated for redundancy at a 99% similarity cutoff. Enrichment was performed using the XGen lockdown protocol and SeqCap EZ Hybridization and Wash Kit (Roche Molecular Systems) according to the manufacturer's instructions.

Using the MAFFT program, the 11 sequences from Bahia (GenBank accession numbers KU0940224, KU0940227, KU0940228, and KX101060-KX101067) were aligned together with all published and available near-complete ZIKV genomes and longer sub-genomic regions (>1500nt) of the Asian genotype as of April of 2016. Both maximum likelihood (ML) and Bayesian phylogenetic inferences were performed using PhyML and BEASTv1.8.2 programs, respectively. The ML phylogeny was reconstructed using the general time reversible nucleotide substitution model with a proportion of invariant sites (GTR+I). Statistical support for phylogenetic nodes was assessed using a bootstrap approach (with 100 replicates). A Bayesian molecular clock phylogeny was estimated using the best fitting evolutionary model determined in Faria, et al (1); specifically, a strict molecular clock GTR+I substitution model, a Bayesian skyline coalescent prior and a non-informative continuous time Markov chain (CTMC) reference prior for the molecular clock rate.

Results

In total, 11 ZIKV partial and complete genome sequences were recovered, with average coverage of 69.4 $\pm$ 2.0% and coverage ranging from 40 - 100% (Table 1). Nearly all isolates from Bahia (11 out of 12) clustered together within a single strongly supported clade (posterior probability 1.00, bootstrap support 96%, Clade C, Figure 1). This support for the Bahia clade is notable given that the majority of ZIKV genomes are incomplete, and this uncertainty is accounted for in the phylogenetic inference. The tree topology obtained is in accordance with previous findings (1,4,6) and the time to the most recent common

ancestor (TRMCA) of the epidemic in the Americas is similar to that previously estimated (1) (Clade A, Figure 1). Our analyses using the new genomes generated here indicate that ZIKV was introduced in Bahia between March and September 2014. An isolate from Maranhão, northeast Brazil, located 1000 km from Bahia, is ancestral to the Bahia clade (bootstrap support 66%, posterior probability 1.00, Clade B, Figure 1). The TRMCA of this northeast Brazil clade (comprising the Bahia clade and the Maranhão sequence) is estimated to be between September 2013 and April 2014, at a very early stage of the epidemic, strengthening the evidence for the hypothesis that ZIKV in the Americas originated from Brazil (1). One previously reported sequence from Bahia (7) clustered (posterior probability=0.99) with an isolate from Belém, Pará state in northern Brazil, which is 3000 km from Bahia. This may represent a separate introduction of ZIKV to Bahia state, as the patient had no history of traveling abroad and denied any sexual contact.

Conclusions

Our results suggest an early introduction and presence (mid-2014) of ZIKV in the state of Bahia, Brazil. The size and statistical support for this cluster make it very likely that this lineage represents a large and sustained chain of transmission within the state. The majority of the cases of the Bahia-specific ZIKV lineage clustered closely to a sequence from Maranhão, and we found evidence for an additional autochthonous transmission event to Bahia from Pará state. Thus, ZIKV in Bahia in mid-2014 is likely to have been introduced from other regions in Brazil rather than from outside the country. The emergence of ZIKV in Bahia state in mid-2014 reported here is consistent with viral infection in pregnant women during the first trimester corresponding to the initial cases of fetal microcephaly reported in Bahia beginning in January 2015 (13), although the peak period of microcephaly did not occur until November of that year.

We cannot yet determine whether the lineage (clade C) identified here comprises the majority of ZIKV cases in Bahia, or, alternatively, whether multiple genetically distinct lineages co-circulate in the state. Brazil currently faces a major public health challenge with co-circulation of ZIKV, DENV, and CHIKV throughout the country (2-4, 14, 15). Additional molecular surveillance in the Americas and beyond is urgently needed to trace and predict the transmission of ZIKV virus.

Acknowledgements

We thank multiple researchers worldwide for permission to include their unpublished ZIKV genomes in our analysis. This study was supported in part by Fundação de Amparo a Pesquisa do Estado da Bahia (SS), the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 614725-PATHPHYLODYN (OGP), National Institutes of Health (NIH) grants R01-HL105704 and R21-AI120977 (CYC), and an award from Abbott Laboratories, Inc. (CYC). 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 - PREDICT. The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the U.S. government.

Conflicts of Interest

C.Y.C. is the director of the UCSF-Abbott Viral Diagnostics and Discovery Center and receives research support from Abbott Laboratories, Inc. and Metabiota, Inc. The other authors disclose no conflicts of interest.

References

1. Musso D, Gubler DJ. Zika Virus. Clinical microbiology reviews. 2016 Jul;29(3):487-524.
2. Faria NR, Azevedo Rdo S, Kraemer MU, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: Early epidemiological and genetic findings. Science. 2016 Apr 15;352(6283):345-9.
3. Campos GS, Bandeira AC, Sardi SI. Zika Virus Outbreak, Bahia, Brazil. Emerging infectious diseases. 2015 Oct;21(10):1885-6.
4. Zanluca C, Melo VC, Mosimann AL, Santos GI, Santos CN, Luz K. First report of autochthonous transmission of Zika virus in Brazil. Memorias do Instituto Oswaldo Cruz. 2015 Jun;110(4):569-72.
5. Secretaria de Vigilância em Saúde. Boletim epidemiológico: Monitoramento dos casos de dengue, febre de chikungunya e febre pelo vírus Zika até a semana epidemiológica 9, 2016 (in Portuguese). Brasilia, Brazil: Brazil Ministry of Health; 2016.
6. Rubin EJ, Greene MF, Baden LR. Zika Virus and Microcephaly. The New England journal of medicine. 2016 Mar 10;374(10):984-5.
7. Giovanetti M, Faria NR, Nunes MR, de Vasconcelos JM, Lourenco J, Rodrigues SG, et al. Zika virus complete genome from Salvador, Bahia, Brazil. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2016 Apr 9;41:142-5.
8. Balm MN, Lee CK, Lee HK, Chiu L, Koay ES, Tang JW. A diagnostic polymerase chain reaction assay for Zika virus. Journal of medical virology. 2012 Sep;84(9):1501-5.
9. Chiu CY, Bres V, Yu G, Krysztof D, Naccache SN, Lee D, et al. Genomic Assays for Identification of Chikungunya Virus in Blood Donors, Puerto Rico, 2014. Emerging infectious diseases. 2015 Aug;21(8):1409-13.
10. Greninger AL, Naccache SN, Messacar K, Clayton A, Yu G, Somasekar S, et al. A novel outbreak enterovirus D68 strain associated with acute flaccid myelitis cases in the USA (2012-14): a retrospective cohort study. The Lancet Infectious diseases. 2015 Jun;15(6):671-82.
11. Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, et al. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. Genome research. 2014 Jul;24(7):1180-92.
12. 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. Emerging infectious diseases. 2008 Aug;14(8):1232-9.
13. Kleber de Oliveira W, Cortez-Escalante J, De Oliveira WT, do Carmo GM, Henriques CM, Coelho GE, et al. Increase in Reported Prevalence of Microcephaly in Infants Born to Women Living in Areas with Confirmed Zika Virus Transmission During the First Trimester of Pregnancy - Brazil, 2015. MMWR Morbidity and mortality weekly report. 2016;65(9):242-7.
14. Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, et al. The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. eLife. 2015;4:e08347.
15. Nunes MR, Faria NR, de Vasconcelos JM, Golding N, Kraemer MU, de Oliveira LF, et al. Emergence and potential for spread of Chikungunya virus in Brazil. BMC medicine. 2015;13:102.

204 **Table 1. Clinical sample information from patients with acute ZIKV infection.**

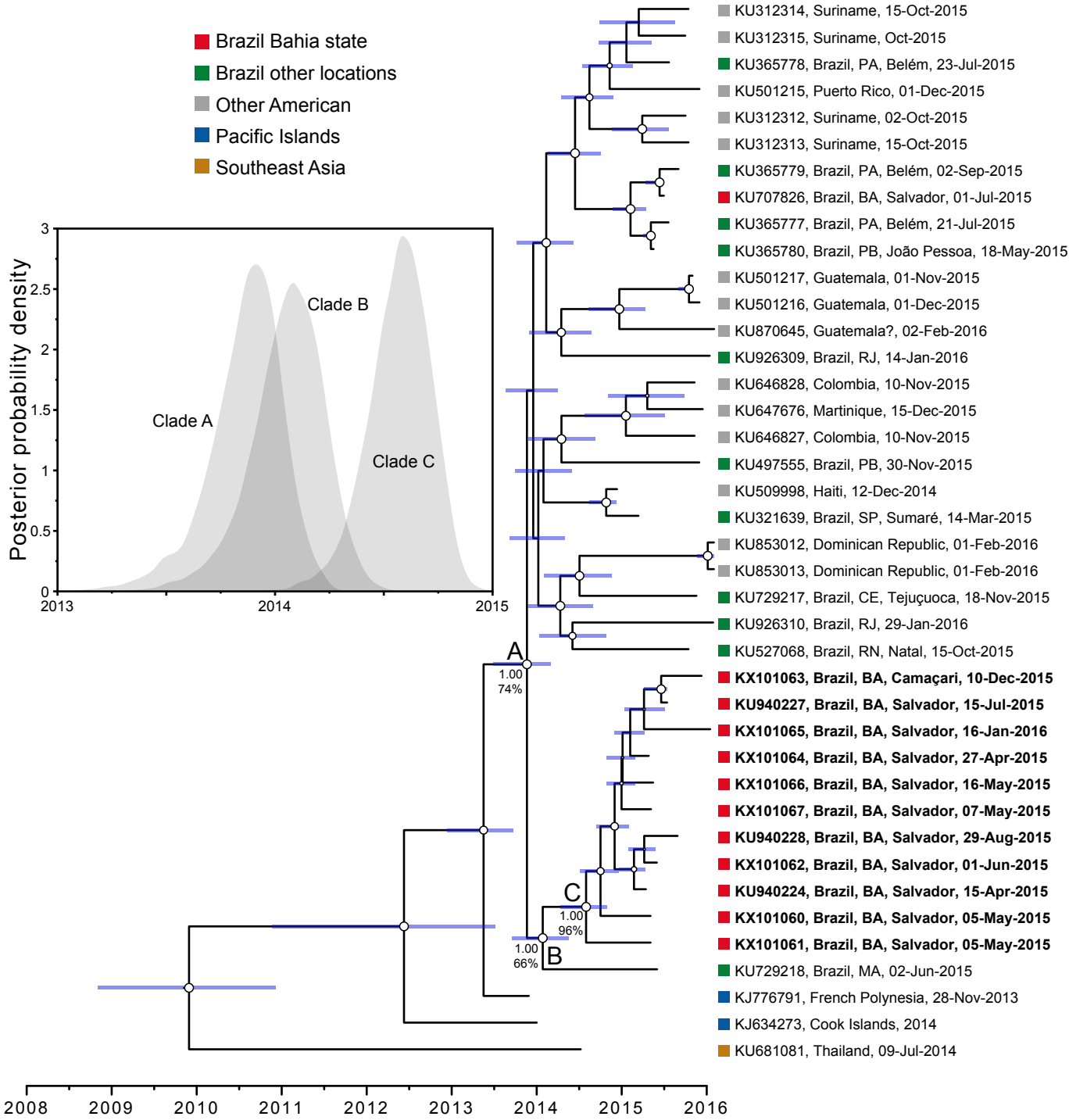
Accession number	Isolate	Patient	Geographical location	ZIKV qRT-PCR cycle threshold	Viral Load (copies / mL)	Collection date	% Genome Recovery (Coverage)
KX101066	Bahia01	Female, 72 years old	Brazil: Bahia, Salvador	34.6	1,042	16-May-2015	84.66
KX101060	Bahia02	Male, 37 years old	Brazil: Bahia, Salvador	32.5	4,086	5-May-2015	74.22
KX101061	Bahia03	Male, 35 years old	Brazil: Bahia, Salvador	32.8	3,272	5-May-2015	77.25
KX101062	Bahia04	Male, 40 years old	Brazil: Bahia, Salvador	34.1	1,464	1-Jun-2015	41.8
KX101063	Bahia05	Male, no data	Brazil: Bahia, Camaçari	33.7	1,901	10-Dec-2015	42.52
KU940228	Bahia07	Female, 37 years old	Brazil: Bahia, Salvador	13.7	9.10E+08	29-Aug-2015	100
KU940227	Bahia08	Female, no data	Brazil: Bahia, Salvador	33.3	2,470	15-Jul-2015	83.74
KU940224	Bahia09	Female, 40 years old	Brazil: Bahia, Salvador	29.9	23,121	25-Apr-2015	99.98
KX101064	Bahia11	Female, 40 years old	Brazil: Bahia, Salvador	–	–	27-Apr-2015	63.61
KX101067	Bahia12	Male, 36 years old	Brazil: Bahia, Salvador	34.2	1,327	7-May-2015	50.85
KX101065	Bahia15	Female, no data	Brazil: Bahia, Salvador	–	–	25-Jan-2016	45.16

205

206

207

Figure 1. Timescale of the ZIKV outbreak in the Americas. A molecular clock phylogeny is shown of the ZIKV outbreak lineage estimated from complete and partial (>1500nt) coding region sequences. For visual clarity, five basal Southeast Asia sequences, HQ23499 (Malaysia, 1966), EU545988 (Micronesia, 2007), KU681082 (Philippines, 2012), JN860885 (Cambodia, 2010) and KU681081 (Thailand, 2013) are not displayed. Blue horizontal bars represent 95% Bayesian credible intervals for divergence dates. A, B and C denote clades discussed in main text and numbers next to them denote posterior probabilities and bootstrap scores (in %). Circle 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 in bold. The inset graph on the left shows the posterior distributions of the estimated ages (TMRCAs) of clades A, B and C, estimated in BEAST using the best fitting evolutionary model.


Brazil Bahia state
Brazil other locations
Other American
Pacific Islands
Southeast Asia
Posterior probability density
Clade A
Clade B
Clade C
0
0.5
1
1.5
2
2.5
3
2013
2014
2015
A
1.00
74%
B
1.00
66%
C
1.00
96%
KU312314, Suriname, 15-Oct-2015
KU312315, Suriname, Oct-2015
KU365778, Brazil, PA, Belém, 23-Jul-2015
KU501215, Puerto Rico, 01-Dec-2015
KU312312, Suriname, 02-Oct-2015
KU312313, Suriname, 15-Oct-2015
KU365779, Brazil, PA, Belém, 02-Sep-2015
KU707826, Brazil, BA, Salvador, 01-Jul-2015
KU365777, Brazil, PA, Belém, 21-Jul-2015
KU365780, Brazil, PB, João Pessoa, 18-May-2015
KU501217, Guatemala, 01-Nov-2015
KU501216, Guatemala, 01-Dec-2015
KU870645, Guatemala?, 02-Feb-2016
KU926309, Brazil, RJ, 14-Jan-2016
KU646828, Colombia, 10-Nov-2015
KU647676, Martinique, 15-Dec-2015
KU646827, Colombia, 10-Nov-2015
KU497555, Brazil, PB, 30-Nov-2015
KU509998, Haiti, 12-Dec-2014
KU321639, Brazil, SP, Sumaré, 14-Mar-2015
KU853012, Dominican Republic, 01-Feb-2016
KU853013, Dominican Republic, 01-Feb-2016
KU729217, Brazil, CE, Tejuçuoca, 18-Nov-2015
KU926310, Brazil, RJ, 29-Jan-2016
KU527068, Brazil, RN, Natal, 15-Oct-2015
KX101063, Brazil, BA, Camaçari, 10-Dec-2015
KU940227, Brazil, BA, Salvador, 15-Jul-2015
KX101065, Brazil, BA, Salvador, 16-Jan-2016
KX101064, Brazil, BA, Salvador, 27-Apr-2015
KX101066, Brazil, BA, Salvador, 16-May-2015
KX101067, Brazil, BA, Salvador, 07-May-2015
KU940228, Brazil, BA, Salvador, 29-Aug-2015
KX101062, Brazil, BA, Salvador, 01-Jun-2015
KU940224, Brazil, BA, Salvador, 15-Apr-2015
KX101060, Brazil, BA, Salvador, 05-May-2015
KX101061, Brazil, BA, Salvador, 05-May-2015
KU729218, Brazil, MA, 02-Jun-2015
KJ776791, French Polynesia, 28-Nov-2013
KJ634273, Cook Islands, 2014
KU681081, Thailand, 09-Jul-2014
2008
2009
2010
2011
2012
2013
2014
2015
2016