

1 **Discovery of a persistent Zika virus lineage in Bahia, Brazil**

2

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

6

7 Affiliations:

8 <sup>1</sup>Department of Laboratory Medicine, University of California, San Francisco, CA 94107,  
9 USA.

10 <sup>2</sup>UCSF-Abbott Viral Diagnostics and Discovery Center, San Francisco, CA 91407, USA.

11 <sup>3</sup>Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

12 <sup>4</sup>Laboratory of Virology, Health Institute of Science, Federal University of Bahia, Salvador,  
13 Bahia, Brazil

14 <sup>5</sup>Department of Laboratory Medicine, University of Washington, Seattle, CA 98195, USA.

15 <sup>6</sup>Alianca Hospital, Salvador, Bahia, Brazil

16 <sup>7</sup>Center of Research Gonzalo Moniz, Foundation Fiocruz, Salvador, Bahia, Brazil

17 <sup>8</sup>Center for Technological Innovation, Evandro Chagas Institute, Ministry of Health,  
18 Ananindeua, PA 67030-000.

19 <sup>9</sup>Department of Medicine, Division of Infectious Diseases, University of California, San  
20 Francisco, CA 94107

21

22 \*Co-first authors

23

24 \*\*Corresponding author: Charles Chiu, MD/PhD, [charles.chiu@ucsf.edu](mailto:charles.chiu@ucsf.edu)

25

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

29

30 Running Title: Persistent ZIKV lineage in Bahia, Brazil

31

## 32 **Abstract**

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

40

## 41 **Introduction**

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

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

52 molecular surveillance in the most affected areas. This is particularly relevant for regions  
53 where other mosquito-borne viruses, particularly dengue virus (DENV) and chikungunya  
54 virus (CHIKV), are known to co-circulate with ZIKV (2), as surveillance based on clinical  
55 symptoms alone can lead to misdiagnosed cases. Accurate genetic characterization of  
56 circulating strains can also facilitate determination of the origin and potential spread of ZIKV  
57 infection in travelers returning from endemic countries. Previous analyses have suggested  
58 that ZIKV was introduced in the Americas at least one year before its initial detection in  
59 Brazil (1). Bahia state comprised 93% of suspected ZIKV cases in Brazil in 2015 (2), and  
60 has reported cases of ZIKV-associated fetal microcephaly, although no genetic information  
61 aside from one complete genome is available from the region to date (2, 7). Here we report  
62 epidemiological findings obtained using 11 new complete and partial ZIKV genomes from  
63 15 clinical cases seen at Hospital Aliança in Salvador, Bahia, between April 2015 and  
64 January 2016.

65

## 66 **Materials and Methods**

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

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

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

95

## 96 **Results**

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

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

116

## 117 **Conclusions**

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

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

135

136

### 137 **Acknowledgements**

138           We thank multiple researchers worldwide for permission to include their  
139 unpublished ZIKV genomes in our analysis. This study was supported in part by Fundação  
140 de Amparo a Pesquisa do Estado da Bahia (SS), the European Research Council under  
141 the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant  
142 agreement no. 614725-PATHPHYLODYN (OGP), National Institutes of Health (NIH) grants  
143 R01-HL105704 and R21-AI120977 (CYC), and an award from Abbott Laboratories, Inc.  
144 (CYC). This study is made possible in part by the generous support of the American people  
145 through the United States Agency for International Development (USAID) Emerging  
146 Pandemic Threats Program - PREDICT. The contents are the responsibility of the authors  
147 and do not necessarily reflect the views of USAID or the U.S. government.

148

149

### 150 **Conflicts of Interest**

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

154

## 155 References

- 156 1. Musso D, Gubler DJ. Zika Virus. *Clinical microbiology reviews*. 2016 Jul;29(3):487-  
157 524.
- 158 2. Faria NR, Azevedo Rdo S, Kraemer MU, Souza R, Cunha MS, Hill SC, et al. Zika  
159 virus in the Americas: Early epidemiological and genetic findings. *Science*. 2016 Apr  
160 15;352(6283):345-9.
- 161 3. Campos GS, Bandeira AC, Sardi SI. Zika Virus Outbreak, Bahia, Brazil. *Emerging*  
162 *infectious diseases*. 2015 Oct;21(10):1885-6.
- 163 4. Zanoluca C, Melo VC, Mosimann AL, Santos GI, Santos CN, Luz K. First report of  
164 autochthonous transmission of Zika virus in Brazil. *Memorias do Instituto Oswaldo Cruz*.  
165 2015 Jun;110(4):569-72.
- 166 5. Secretaria de Vigilância em Saúde. Boletim epidemiológico: Monitoramento dos  
167 casos de dengue, febre de chikungunya e febre pelo vírus Zika até a semana  
168 epidemiológica 9, 2016 (in Portuguese). Brasilia, Brazil: Brazil Ministry of Health; 2016.
- 169 6. Rubin EJ, Greene MF, Baden LR. Zika Virus and Microcephaly. *The New England*  
170 *journal of medicine*. 2016 Mar 10;374(10):984-5.
- 171 7. Giovanetti M, Faria NR, Nunes MR, de Vasconcelos JM, Lourenco J, Rodrigues  
172 SG, et al. Zika virus complete genome from Salvador, Bahia, Brazil. *Infection, genetics and*  
173 *evolution : journal of molecular epidemiology and evolutionary genetics in infectious*  
174 *diseases*. 2016 Apr 9;41:142-5.
- 175 8. Balm MN, Lee CK, Lee HK, Chiu L, Koay ES, Tang JW. A diagnostic polymerase  
176 chain reaction assay for Zika virus. *Journal of medical virology*. 2012 Sep;84(9):1501-5.
- 177 9. Chiu CY, Bres V, Yu G, Kryzstof D, Naccache SN, Lee D, et al. Genomic Assays for  
178 Identification of Chikungunya Virus in Blood Donors, Puerto Rico, 2014. *Emerging*  
179 *infectious diseases*. 2015 Aug;21(8):1409-13.
- 180 10. Greninger AL, Naccache SN, Messacar K, Clayton A, Yu G, Somasekar S, et al. A  
181 novel outbreak enterovirus D68 strain associated with acute flaccid myelitis cases in the  
182 USA (2012-14): a retrospective cohort study. *The Lancet Infectious diseases*. 2015  
183 Jun;15(6):671-82.
- 184 11. Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, et  
185 al. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from  
186 next-generation sequencing of clinical samples. *Genome research*. 2014 Jul;24(7):1180-92.
- 187 12. Lanciotti RS, Kosoy OL, Laven JJ, Velez JO, Lambert AJ, Johnson AJ, et al.  
188 Genetic and serologic properties of Zika virus associated with an epidemic, Yap State,  
189 Micronesia, 2007. *Emerging infectious diseases*. 2008 Aug;14(8):1232-9.
- 190 13. Kleber de Oliveira W, Cortez-Escalante J, De Oliveira WT, do Carmo GM,  
191 Henriques CM, Coelho GE, et al. Increase in Reported Prevalence of Microcephaly in  
192 Infants Born to Women Living in Areas with Confirmed Zika Virus Transmission During the  
193 First Trimester of Pregnancy - Brazil, 2015. *MMWR Morbidity and mortality weekly report*.  
194 2016;65(9):242-7.
- 195 14. Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, et al. The  
196 global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *eLife*.  
197 2015;4:e08347.
- 198 15. Nunes MR, Faria NR, de Vasconcelos JM, Golding N, Kraemer MU, de Oliveira LF,  
199 et al. Emergence and potential for spread of Chikungunya virus in Brazil. *BMC medicine*.  
200 2015;13:102.

201  
202  
203

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



208 **Figure 1. Timescale of the ZIKV outbreak in the Americas.** A molecular clock phylogeny  
209 is shown of the ZIKV outbreak lineage estimated from complete and partial (>1500nt)  
210 coding region sequences. For visual clarity, five basal Southeast Asia sequences,  
211 HQ23499 (Malaysia, 1966), EU545988 (Micronesia, 2007), KU681082 (Philippines, 2012),  
212 JN860885 (Cambodia, 2010) and KU681081 (Thailand, 2013) are not displayed. Blue  
213 horizontal bars represent 95% Bayesian credible intervals for divergence dates. A, B and C  
214 denote clades discussed in main text and numbers next to them denote posterior  
215 probabilities and bootstrap scores (in %). Circle sizes represent, at each node, the posterior  
216 probability support of that node. Taxa are labeled with accession number, sampling  
217 location, and sampling date. Names of sequences generated in this study are in bold. The  
218 inset graph on the left shows the posterior distributions of the estimated ages (TMRCA) of  
219 clades A, B and C, estimated in BEAST using the best fitting evolutionary model.

220

221

222

223

- Brazil Bahia state
- Brazil other locations
- Other American
- Pacific Islands
- Southeast Asia

