

# Genome Sequence of a Candidate World Health Organization Reference Strain of Zika Virus for Nucleic Acid Testing

Jan-Hendrik Trösemeier,<sup>a,b</sup> Didier Musso,<sup>c</sup> Johannes Blümel,<sup>a</sup> Julien Thézè,<sup>d</sup> Oliver G. Pybus,<sup>d</sup> Sally A. Baylis<sup>a</sup>

Paul-Ehrlich-Institut, Langen, Germany<sup>a</sup>; Molecular Bioinformatics, Johann Wolfgang Goethe University, Frankfurt am Main, Germany<sup>b</sup>; Institut Louis Malardé, Tahiti, French Polynesia<sup>c</sup>; Department of Zoology, University of Oxford, Oxford, United Kingdom<sup>d</sup>

**We report here the sequence of a candidate reference strain of Zika virus (ZIKV) developed on behalf of the World Health Organization (WHO). The ZIKV reference strain is intended for use in nucleic acid amplification (NAT)-based assays for the detection and quantification of ZIKV RNA.**

Received 6 July 2016 Accepted 6 July 2016 Published 1 September 2016

**Citation** Trösemeier J-H, Musso D, Blümel J, Thézè J, Pybus OG, Baylis SA. 2016. Genome sequence of a candidate World Health Organization reference strain of Zika virus for nucleic acid testing. *Genome Announc* 4(5):00917-16. doi:10.1128/genomeA.00917-16.

**Copyright** © 2016 Trösemeier et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Sally A. Baylis, [sally.baylis@pei.de](mailto:sally.baylis@pei.de).

The mosquito-borne Zika virus (ZIKV) is a member of the *Flavivirus* genus in the *Flaviviridae* family, and nearly 70 years after its discovery in Uganda, it is spreading widely in South and Central America, the Caribbean, and the Pacific. Infections are frequently asymptomatic or cause mild disease; however, severe complications have recently come to light, including central nervous system abnormalities (especially microcephaly) in fetuses/neonates and Guillain-Barré syndromes in adults. ZIKV is transmitted by the bite of infected mosquitoes, but nonvector-borne transmission has been reported, through materno/fetal, sexual, and blood transfusion (1–3). Diagnosis of acute ZIKV infection relies upon the detection of virus RNA in plasma, urine, and other body fluids (2, 4). In order to implement robust diagnostic testing methods, a candidate World Health Organization (WHO) reference material (international standard), intended to be used for the quantitation of ZIKV RNA and evaluated in an international study, was developed. The intended WHO ZIKV reference material is similar to those developed for other blood-borne viruses (5).

The ZIKV strain PF13/251013-18 was isolated from serum specimen from a French Polynesia patient in 2013 (6). ZIKV RNA was extracted from Vero E6 cell culture supernatants, from low-passage virus, using the QIAamp MinElute virus spin kit (Qiagen GmbH, Hilden, Germany). The passage number was identical to the virus stock used in the preparation of the lyophilized candidate reference material. Sequencing was performed using Illumina MiSeq (2 × 300 bp).

Sequencing reads were preprocessed by quality filtering with PRINSEQ (7). *De novo* assembly was performed with the MIRA 4 assembler (8). On the basis of BLAST searching against the NCBI database, the sequence with the highest identity and *E* score was chosen for reference-assisted scaffolding using Mira and BWA (9). Consensus sequences, based on reference-assisted scaffolding, were constructed using SAMtools (10). The final ZIKV sequence has a length of 10,769 bp.

The ZIKV strain PF13/251013-18 genome sequence was aligned along with all published complete and partial (>80% coverage) ZIKV genomes belonging to the Asian genotype using

MAFFT (11). A maximum-likelihood phylogenetic tree was inferred using PhyML (12) under a reversible nucleotide substitution model with a proportion of invariant sites (GTR + I), which was the best-fitting model, as determined by jModelTest (13). Statistical support for phylogenetic nodes was assessed using a bootstrap approach (with 100 replicates). The resulting ZIKV Asian genotype phylogeny is in accordance with previous studies (14). ZIKV strain PF13/251013-18 is an outgroup to the American clade (bootstrap support, 77%) and is genetically similar to another genome from the French Polynesia outbreak (isolate H/PF/2013).

The demonstration that the WHO ZIKV reference strain is closely related to currently circulating viruses in the Pacific, Central and South America, and Caribbean is important, because the selection of viruses for the development of WHO reference materials has been based on strains representative of viruses that are clinically significant and have widespread distribution. Knowledge of the sequence of the PF13/251013-18 ZIKV isolate is essential for oligonucleotide design for NAT assays to ensure adequate detection and quantification of the WHO reference material.

**Accession number(s).** The sequence of the candidate WHO reference strain PF13/251013-18 has been deposited in GenBank under accession no. [KX369547](https://www.ncbi.nlm.nih.gov/nuclot/KX369547).

## ACKNOWLEDGMENT

We thank Roswitta Kleiber for assistance.

## FUNDING INFORMATION

This work was supported by the grant "Sicherheit von Blut(produkten) und Geweben hinsichtlich der Abwesenheit von Zikaviren" from the German Ministry of Health (S. A. Baylis, J.-H. Trösemeier, J. Blümel). This study was made possible by the generous support of the American people through the U.S. Agency for International Development (USAID) Emerging Pandemic Threats Program-2 PREDICT-2 (cooperative agreement no. AID-OAA-A-14-00102; J. Thézè, O. Pybus). The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the U.S. government.

## REFERENCES

1. Cunha MS, Esposito DL, Rocco IM, Maeda AY, Vasami FG, Nogueira JS, de Souza RP, Suzuki A, Addas-Carvalho M, Barjas-Castro Mde L, Resende MR, Stucchi RS, Boin Ide F, Katz G, Angerami RN, da Fonseca BA. 2016. First complete genome sequence of Zika virus (*Flaviviridae*, *Flavivirus*) from an autochthonous transmission in Brazil. *Genome Announc* 4(2):e00032-16. <http://dx.doi.org/10.1128/genomeA.00032-16>.
2. Musso D, Gubler DJ. 2016. Zika virus. *Clin Microbiol Rev* 29:487–524. <http://dx.doi.org/10.1128/CMR.00072-15>.
3. Musso D, Stramer SL, Busch MP. 2016. Zika virus: a new challenge for blood transfusion. *Lancet* 387:1993–1994. [http://dx.doi.org/10.1016/S0140-6736\(16\)30428-7](http://dx.doi.org/10.1016/S0140-6736(16)30428-7).
4. Centers for Disease Control and Prevention. Interim guidance for Zika virus testing—United States, 2016. *MMWR Morb Mortal Wkly Rep* 65:474.
5. Baylis SA, Chudy M, Nübling CM. 2015. Standardization of NAT for blood-borne pathogens. *Transfus Med Hemother* 42:211–218. <http://dx.doi.org/10.1159/000435872>.
6. Aubry M, Richard V, Green J, Broult J, Musso D. 2016. Inactivation of Zika virus in plasma with amotosalen and ultraviolet A illumination. *Transfusion* 56:33–40. <http://dx.doi.org/10.1111/trf.13271>.
7. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27:863–864. <http://dx.doi.org/10.1093/bioinformatics/btr026>.
8. Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. *In* Computer science and biology. Proceedings of the German Conference on Bioinformatics, GCB '99. GCB, Hannover, Germany.
9. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595. <http://dx.doi.org/10.1093/bioinformatics/btp698>.
10. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <http://dx.doi.org/10.1093/bioinformatics/btp352>.
11. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <http://dx.doi.org/10.1093/molbev/mst010>.
12. Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704. <http://dx.doi.org/10.1080/10635150390235520>.
13. Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772. <http://dx.doi.org/10.1038/nmeth.2109>.
14. Faria NR, Azevedo Rdo S, Kraemer MU, Souza R, Cunha MS, Hill SC, Théze J, Bonsall MB, Bowden TA, Rissanen I, Rocco I, Nogueira JS, Maeda AY, Vasami FG, Macedo FL, Suzuki A, Rodrigues SG, Cruz AC, Nunes BT, Medeiros DB, Rodrigues DS, Nunes Queiroz AL, da Silva EV, Henriques DF, Travassos da Rosa ES, de Oliveira CS, Martins LC, Vasconcelos HB, Casseb LM, Simith Dde B, Messina JP, Abade L, Lourenço J, Carlos Junior Alcantara L, de Lima MM, Giovanetti M, Hay SI, de Oliveira RS, Lemos Pda S, de Oliveira LF, de Lima CP, da Silva SP, de Vasconcelos JM, Franco L, Cardoso JF, Vianez-Júnior JL, Mir D, Bello G, Delatorre E, et al. 2016. Zika virus in the Americas: early epidemiological and genetic findings. *Science* 352:345–349. <http://dx.doi.org/10.1126/science.aaf5036>.