

# Complete Genome Sequences of Five Zika Virus Isolates

Jason T. Ladner,<sup>a</sup> Michael R. Wiley,<sup>a</sup> Karla Prieto,<sup>a</sup> Chadwick Y. Yasuda,<sup>b\*</sup> Elyse Nagle,<sup>a</sup> Matthew R. Kasper,<sup>c</sup> Daniel Reyes,<sup>a</sup> Nikolaos Vasilakis,<sup>d,e,f</sup> Vireak Heang,<sup>b</sup> Scott C. Weaver,<sup>d,e,f</sup> Andrew Haddow,<sup>g</sup> Robert B. Tesh,<sup>d,e,f</sup> Ly Sovann,<sup>h</sup> Gustavo Palacios<sup>a</sup>

Center for Genome Sciences, U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, Maryland, USA<sup>a</sup>; U.S. Naval Medical Research Unit No.2, Det Phnom Penh (NAMRU-2), Cambodia<sup>b</sup>; Naval Medical Research Center, Silver Spring, Maryland, USA<sup>c</sup>; Institute for Human Infections and Immunity, University of Texas Medical Branch, Galveston, Texas, USA<sup>d</sup>; Department of Pathology, University of Texas Medical Branch, Galveston, Texas, USA<sup>e</sup>; Department of Microbiology and Immunology, University of Texas Medical Branch, Galveston, Texas, USA<sup>f</sup>; Division of Virology, U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, Maryland, USA<sup>g</sup>; Communicable Disease Centre, Cambodian Ministry of Health, Phnom Penh, Cambodia<sup>h</sup>

\* Present address: Chadwick Y. Yasuda, U.S. Naval Research Laboratory, Washington, DC, USA.

**Zika virus is an emerging human pathogen of great concern due to putative links to microcephaly and Guillain-Barre syndrome. Here, we report the complete genomes, including the 5' and 3' untranslated regions, of five Zika virus isolates, one from the Asian lineage and four from the African lineage.**

Received 23 March 2016 Accepted 24 March 2016 Published 12 May 2016

**Citation** Ladner JT, Wiley MR, Prieto K, Yasuda CY, Nagle E, Kasper MR, Reyes D, Vasilakis N, Heang V, Weaver SC, Haddow A, Tesh RB, Sovann L, Palacios G. 2016. Complete genome sequences of five Zika virus isolates. *Genome Announc* 4(3):e00377-16. doi:10.1128/genomeA.00377-16.

**Copyright** © 2016 Ladner 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 Gustavo Palacios, [gustavo.f.palacios.ctr@mail.mil](mailto:gustavo.f.palacios.ctr@mail.mil).

Zika virus (ZIKV) is an emerging human pathogen of great concern due to putative links to severe birth defects (microcephaly [1]) and a rare autoimmune syndrome (Guillain-Barre [2]). The virus was first detected in a sentinel rhesus macaque in 1947 while monitoring for enzootic yellow fever virus circulation in the Zika Forest, Uganda (3). Since then, ZIKV has been detected throughout much of Sub-Saharan Africa and has spread across the Pacific Ocean to Central and South America via Southeast Asia (4, 5).

ZIKV is an arbovirus believed to be vectored mainly by mosquitoes in the genus *Aedes*. ZIKV belongs to the Spondweni serocomplex within the genus *Flavivirus* and the family *Flaviviridae*. Viruses in the genus *Flavivirus* have single-stranded positive-sense RNA genomes of ~11 kb in size (6). These genomes encode a single polyprotein flanked by 5' and 3' untranslated regions (UTRs).

Here, we report five complete ZIKV genomes, including the full polyprotein and the 5' and 3' UTRs. Four of these viruses were isolated in Africa (Senegal and Uganda) and belong to the African lineage (4). One virus (FSS13025) was isolated in Cambodia and belongs to the Asian lineage (4). All five isolates were passaged multiple times in cell culture (Table 1). Several distinctly passaged versions of strain MR-766 have been previously published, includ-

ing two complete genomes (GenBank: AY632535, LC002520). A coding-complete version (i.e., missing the 5' and 3' UTRs [7]) of the genome was published for an earlier passage of FSS13025 (GenBank: JN860885).

Isolates were sequenced at the U.S. Army Medical Research Institute of Infectious Diseases and the University of Texas Medical Branch using Illumina NextSeq/MiSeq. RNA was extracted using the Direct-zol RNA extraction kit (Zymo), converted to cDNA using SuperScript III (Invitrogen) and amplified using sequence-independent single-primer amplification (8) combined with primers for rapid amplification of cDNA ends (9). Adaptors and primers were clipped using Cutadapt version 1.21 (10), and low-quality reads/bases were filtered using Prinseq-lite version 0.20.4 (11). Reads were aligned to a reference genome (chosen separately for each isolate based on nucleotide-level similarity between preliminary *de novo* contigs and publically available ZIKV genomes) using Bowtie2 (12), duplicates were removed with Picard (<http://broadinstitute.github.io/picard>), and a new consensus was generated using a combination of SAMtools version 0.1.18 (13) and custom scripts. When necessary, the 5' and 3' UTRs were then built out beyond the reference genome using custom scripts that iterate the reference assembly process with the

**TABLE 1** Accession numbers and source information for sequenced Zika virus isolates

Isolate	Source	Location	Date	Passage history <sup>a</sup>	Genome/ORF <sup>b</sup> size (nucleotides)	Accession no.
FSS13025	<i>Homo sapiens</i>	Cambodia	8/8/2010	Vero ×4	10,807/10,272	KU955593
DAK AR D 41671	<i>Aedes taylori</i>	Senegal	12/14/1984	AP61 ×1, C6/36 ×1, Vero ×3	10,806/10,272	KU955595
DAK AR D 41662	<i>Aedes taylori</i>	Senegal	12/6/1984	AP61 ×1, C6/36 ×1, Vero ×3	10,806/10,272	KU955592
DAK AR D 41525	<i>Aedes africanus</i>	Senegal	11/20/1984	AP61 ×1, C6/36 ×1, Vero ×3	10,806/10,272	KU955591
MR-766	<i>Macaca mulatta</i>	Uganda	4/1947	SM ×150, Vero ×2	10,795/10,260	KU955594

<sup>a</sup> AP61, *Aedes pseudoscutellaris* cells; C6/36, *Aedes albopictus* cells; SM, suckling mouse; Vero, African green monkey kidney cells.

<sup>b</sup> ORF, open reading frame.

addition of ambiguous characters (N's) onto the 5' and 3' ends of the reference.

The assembled, complete genomes (7) ranged in size from 10,795 to 10,807 nucleotides; the 5' UTRs were 106 to 107 nucleotides and the 3' UTRs were 428 to 429 nucleotides, respectively. The MR-766 stock we sequenced contained a 12-nucleotide in-frame deletion in the polyprotein, which it shared with AY632535, but is not present in LC002520. However, our sequence does not share the five additional single nucleotide insertions/deletions present in AY632535. All sequenced isolates contained the conserved, dinucleotide complimentary terminal sequences that are characteristic of the *Flavivirus* genus (5' – AG . . . CU – 3' [6]).

**Nucleotide sequence accession numbers.** Genome accession numbers to public databases are listed in Table 1.

## ACKNOWLEDGMENTS

We thank Bradley Pfeffer for his help submitting the genomes to GenBank. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the U.S. Army. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government. Some authors of this work are military service members or employees of the U.S. Government. This work was prepared as part of their official duties.

The work at USAMRIID was funded by DTRA project CB10246. Some of the genomes referenced in this publication were also sequenced through funding provided by DHS S&T through contract no. HSHQDC-13-C-B0009 “Capturing Global Biodiversity of Pathogens by Whole Genome Sequencing.”

## FUNDING INFORMATION

This work, including the efforts of Nikolaos Vasilakis and Scott C. Weaver, was funded by Department of Homeland Security Science and Technology (HSHQDC-13-C-B0009). This work, including the efforts of Jason T. Ladner, Michael R. Wiley, Karla Prieto, Elyse Nagle, Daniel Reyes, and Gustavo Palacios, was funded by DOD | Defense Threat Reduction Agency (DTRA) (CB10246).

## REFERENCES

1. Mlakar J, Korva M, Tul N, Popović M, Poljšak-Prijatelj M, Mraz J, Kolenc M, Resman Rus K, Vesnaver Vipotnik T, Fabjan Vodusek V, Vizjak A, Pizem J, Petrovec M, Avšič Županc T. Zika virus associated

- with microcephaly. *N Engl J Med* 374:951–958. <http://dx.doi.org/10.1056/NEJMoa1600651>.
2. Oehler E, Watrin L, Larre P, Leparc-Goffart I, Lastere S, Valour F, Baudouin L, Mallet H, Musso D, Ghawche F. 2014. Zika virus infection complicated by Guillain-Barre syndrome—case report, French Polynesia, December 2013. *Euro Surveill* 19:20720. <http://dx.doi.org/10.2807/1560-7917.ES2014.19.9.20720>.
3. Dick GWA, Kitchen SF, Haddock AJ. 1952. Zika virus (I). Isolations and serological specificity. *Transactions R Soc Trop Med Hyg* 46:509–520. [http://dx.doi.org/10.1016/0035-9203\(52\)90042-4](http://dx.doi.org/10.1016/0035-9203(52)90042-4).
4. Haddock AD, Schuh AJ, Yasuda CY, Kasper MR, Heang V, Huy R, Guzman H, Tesh RB, Weaver SC. 2012. Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. *PLoS Negl Trop Dis* 6:e1477. <http://dx.doi.org/10.1371/journal.pntd.0001477>.
5. Ios S, Mallet H-P, Leparc Goffart I, Gauthier V, Cardoso T, Herida M. 2014. Current Zika virus epidemiology and recent epidemics. *Med Mal Infectieuses* 44:302–307. <http://dx.doi.org/10.1016/j.medmal.2014.04.008>.
6. King AM, Adams MJ, Lefkowitz EJ. 2011. Virus taxonomy: classification and nomenclature of viruses: ninth report of the International Committee on Taxonomy of Viruses, vol 9. Academic Press, London.
7. Ladner JT, Beitzel B, Chain PS, Davenport MG, Donaldson EF, Friedman M, Kugelman JR, Kuhn JH, O'Rear J, Sabeti PC, Wentworth DE, Wiley MR, Yu GY, Threat Characterization Consortium, Sozhamannan S, Bradburne C, Palacios G. 2014. Standards for sequencing viral genomes in the era of high-throughput sequencing. *mBio* 5:e01360-01314. <http://dx.doi.org/10.1128/mBio.01360-14>.
8. Djikeng A, Halpin R, Kuzmickas R, Depasse J, Feldblyum J, Sengamalay N, Afonso C, Zhang X, Anderson NG, Ghedin E, Spiro DJ. 2008. Viral genome sequencing by random priming methods. *BMC Genomics* 9:5. <http://dx.doi.org/10.1186/1471-2164-9-5>.
9. Leguia M, Loyola S, Rios J, Juarez D, Guevara C, Silva M, Prieto K, Wiley M, Kasper MR, Palacios G, Bausch DG. 2015. Full genomic characterization of a Saffold virus isolated in Peru. *Pathogens* 4:816–825. <http://dx.doi.org/10.3390/pathogens4040816>.
10. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10–12. <http://dx.doi.org/10.14806/ej.17.1.200>.
11. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27:863–864. <http://dx.doi.org/10.1093/bioinformatics/btr026>.
12. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <http://dx.doi.org/10.1038/nmeth.1923>.
13. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <http://dx.doi.org/10.1093/bioinformatics/btp352>.