

## LETTER TO THE EDITOR

# From discovery to outbreak: the genetic evolution of the emerging Zika virus

Hong Liu<sup>1,\*</sup>, Liang Shen<sup>1,\*</sup>, Xiao-Lin Zhang<sup>1</sup>, Xiao-Long Li<sup>2,3</sup>, Guo-Dong Liang<sup>2,3</sup> and Hong-Fang Ji<sup>1</sup>*Emerging Microbes & Infections* (2016) 5, e111; doi:10.1038/emi.2016.109; published online 26 October 2016**Dear Editor,**

Zika virus (ZIKV) was initially isolated from a rhesus monkey in Uganda in 1947.<sup>1</sup> Thereafter, ZIKV was identified in both mosquitoes and humans, and mosquitoes are the transmitting vector of ZIKV.<sup>2</sup> Historically, only a few infectious cases were reported worldwide before the 21st century, but in 2007, a large ZIKV epidemic occurred on Yap island.<sup>3</sup> Another outbreak is currently spreading throughout South America, resulting in a large number of cases of microcephaly,<sup>4</sup> Guillain–Barre syndrome<sup>5</sup> and viral meningoencephalitis.<sup>6</sup> Unlike other arboviruses, ZIKV can be transmitted through sexual contact. Thus, the characteristics of ZIKV infection are more complex and diverse than those of any other arboviral disease. For these reasons, the World Health Organization (WHO) has declared the present ZIKV outbreak as an international public health issue.<sup>7</sup>

A phylogenetic analysis of ZIKV sequences identified two major lineages: the African and Asian lineages.<sup>8,9</sup> Asian strains of ZIKV are responsible for the outbreaks occurring in South America since 2015. Since the beginning of the 21st century, a significant increase in the number of ZIKV cases has been reported.<sup>2</sup> Although ZIKV has been circulating in nature for almost 70 years, the reason for the sudden increase in the virulence and transmission capacity of the virus is unknown. The underlying mechanisms of ZIKV molecular evolution are urgently in need of clarification. In this study, we used Bayesian methods in a phylogenetic analysis of whole-genome sequences from the currently circulating ZIKV strains to determine the genetic evolutionary trends of ZIKV and provide some clues for ZIKV prevention and control.

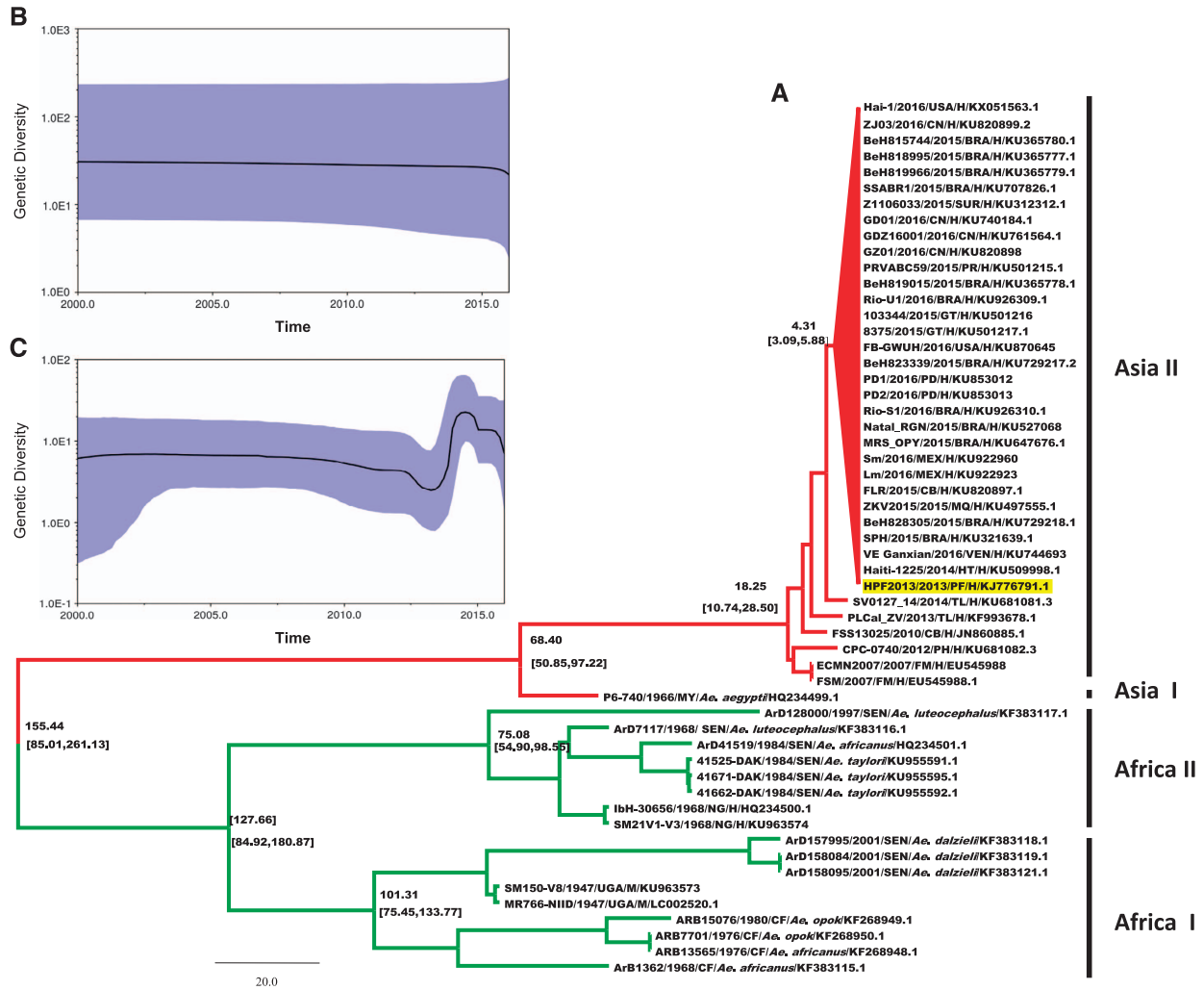
The complete genome sequences from a total of 55 ZIKV isolates were obtained from the GenBank database (as of 1 May 2016). These strains were isolated from a variety of hosts: monkeys ( $n=2$ ), mosquitoes ( $n=14$ ) and humans ( $n=39$ ). The isolation dates range from 1947 to 2016, and the locations of isolation include all regions in which ZIKV cases have been identified (Figure 1A). Bayesian time-scaled phylogenetic analysis, analysis of the molecular evolutionary rates of the entire ZIKV population and of each ZIKV genotype, and the time to the most recent common ancestor (TMRCA) of ZIKV

were estimated comprehensively using the BEAST software package.<sup>10</sup> The GTR+I+G substitution model was selected as the best-fit nucleotide substitution model by MrModelTest. The relaxed log clock model and Bayesian skyline plot demographics were chosen as the best-fit parameters. The analysis was run through 50 000 000 generations to ensure sufficient mixing. Finally, the maximum clade credibility (MCC) tree was built using TreeAnnotator with 10% burn-in (<http://beast.bio.ed.ac.uk/>). The MCC tree based on the whole genome of each ZIKV strain is shown in Figure 1A. The posterior probability values for the nodes of each lineage were  $>0.8$ , indicating their robustness. The most recent common ancestor for all ZIKV genotypes was estimated to have arisen ~155 years ago (95% highest posterior density (HPD): 85.01–261.13), and this common ancestor subsequently evolved into two main populations: the African and Asian lineages. The TMRCA for the Africa lineage was estimated at 127 years ago (95% HPD: 84.92–180.87), while that of the Asian lineage was estimated at 68 years ago (95% HPD: 50.85–97.22). The African lineage is divided into two independent evolutionary groups: African sublineages I and II. The TMRCA for African sublineage I, which comprises two isolates from monkeys and nine from mosquitoes, was estimated at 101 years ago (95% HPD: 75.45–133.77). The TMRCA for African sublineage II, which includes two strains isolated from patients in 1968, was estimated at 75 years ago (95% HPD: 54.90–98.55). The Asian lineage could also be divided into two groups based on the host species: Asian sublineages I and II. For Asian sublineage I, the single isolate from a mosquito (P6-740/Malaysia/1966) forms an independent evolutionary branch and is the oldest member of the Asian lineage. All of the human strains isolated from 2007 to 2016 cluster together to form Asian sublineage II. This sublineage emerged 18 years ago (95% HPD: 10.74–28.50) and is the youngest lineage within the entire ZIKV population. All of the strains isolated from the epidemic in 2015 to 2016, together with the Haiti225 (Haiti, 2014) isolate and the H/PF2013 (French Polynesia, 2013) isolate, exhibited a close relationship and formed a group. Moreover, the H/PF2013 strain is located close to the root of the phylogenetic group, indicating that it is likely the ancestor of the other strains in

<sup>1</sup>Shandong Provincial Research Center for Bioinformatic Engineering and Technique, School of Life Sciences, Shandong University of Technology, Zibo 255049, Shandong Province, China; <sup>2</sup>State Key Laboratory of Infectious Disease Prevention and Control, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China and <sup>3</sup>Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Hangzhou 310058, Zhejiang Province, China  
Correspondence: GD Liang; HF Ji  
E-mail: gdliang@hotmail.com; jhf@sdut.edu.cn

\*These authors contributed equally to this work.

Received 24 June 2016; revised 24 August 2016; accepted 25 August 2016



**Figure 1** Maximum clade credibility tree and Bayesian skyline plots for African and Asian lineages of ZIKV. **(A)** Maximum clade credibility tree for 55 whole-genome sequences of ZIKV obtained from isolates collected in 1947 to 2016. Consistent with previous studies, the tree is divided into two distinct lineages: the African lineage (green) and the Asian lineage (red). Estimated TMRCA of these lineages (with their 95% HPD values in parentheses) are shown. The current pandemic ZIKV group of Asian lineage is shown as a red triangle, and strain H/PPF2013 is highlighted in yellow. Viral nomenclature is as follows: virus strain/year of isolation/country/origin/accession number. BRA, Brazil; CB, Cambodia; CN, China; FM, Micronesia; GF, Central African Republic; GT, Guatemala; H, Human; HT, Haiti; M, Monkey; MEX, Mexico; MQ, Martinique; MY, Malaysia; NG, Nigeria; PD, Padua; PF, French Polynesia; PH, Philippines; PR, Puerto Rico; SEN, Senegal; SUR, Suriname; TL, Thailand; UGA, Uganda; VEN, Venezuela. **(B, C)** Bayesian skyline plots for African and Asian lineages, separately, since the beginning of the 21st century. The horizontal and vertical axes represent time and genetic diversity, respectively. The highlighted areas correspond to 95% HPD intervals.

this population. Furthermore, the TMRCA for this group was estimated at 4.31 years ago (95% HPD: 3.09–5.88), suggesting that the ancestor of this sublineage was circulating in 2012.

The TMRCA for all ZIKV strains was estimated to be 155 years ago in our current study, while the TMRCA for all Japanese encephalitis virus strains<sup>11</sup> was estimated to diverge approximately 1695 years ago and those for West Nile virus<sup>12</sup> and yellow fever virus<sup>13</sup> at approximately 200 years ago and 306 years ago, respectively. Thus, ZIKV can be considered an ‘emerging’ arbovirus. Substitution rate is an important parameter for understanding viral evolution. In this study, the mean nucleotide substitution rate for the entire ZIKV population was estimated at  $6.091 \times 10^{-4}$  substitutions per site per year (s/s/y) (95% HPD,  $3.44 \times 10^{-4}$  to  $8.79 \times 10^{-4}$ ). Furthermore, the estimated substitution rate for the African lineage was  $2.689 \times 10^{-4}$  s/s/y (95% HPD,

$6.241 \times 10^{-6}$  to  $6.306 \times 10^{-4}$ ), while it was  $1.0233 \times 10^{-3}$  s/s/y (95% HPD,  $8.2 \times 10^{-4}$  to  $1.2 \times 10^{-3}$ ) for the Asian lineage. Surprisingly, the substitution rate of the currently circulating Asian lineage of ZIKV is considerably higher than those of the African lineage (pre-pandemic ZIKV population) and other mosquito-borne flaviviruses, all of which have substitution rates at  $10^{-4}$  s/s/y. Considering that a higher substitution rate often enhances viral adaptability and competitiveness, as well as pathogenicity,<sup>14</sup> this rate could be a contributing factor in the sudden outbreak of ZIKV caused by strains of the Asian lineage.

According to our results, since the marked increase in cases of ZIKV infection in the 21st century, obvious differences have arisen between the African and Asian lineages with respect to genetic diversity. As shown in Figure 1B, the genetic diversity of the African lineage has remained relatively stable over time, while the Asian lineage

(Figure 1C) underwent more active changes and maintained a relatively higher level of diversity than that of the 2013 ZIKV outbreak in French Polynesia.

Our work provides further insight regarding the molecular evolutionary differences between the African and Asian lineages of ZIKV. The Asian lineage, which is the youngest genotype, has a fast rate of evolution and a marked increase in genetic diversity, which provides this lineage with a selective advantage. Notably, a recent report by the WHO on the spread of the Asian lineage of ZIKV beyond South America to the country of Cape Verde in West Africa<sup>7</sup> indicates the potential for ZIKV to spread to non-endemic areas. The ZIKV infections that occurred in 2015 were mainly distributed in tropical countries such as Brazil in South America, but countries in the northern hemisphere were not conducive for mosquito breeding in the winter. However, with the northern hemisphere gradually entering into the summer season, with increasing temperature, precipitation and mosquito density, ZIKV will likely spread northward. Therefore, enhanced detection and surveillance of ZIKV in the northern hemisphere is highly recommended despite the current pandemic being restricted to the southern hemisphere.

#### ACKNOWLEDGEMENTS

This work was supported by grants from the National Natural Science Foundation of China (81290342 and 81301479), the Shandong Provincial Science Foundation for Distinguished Young Scholars (JQ201508), the Key Program of the Shandong Provincial Science Foundation (ZR2015JL010), a Development Grant of the State Key Laboratory of Infectious Disease Prevention and Control (2014SKLID103), the Natural Science Foundation of Shandong Province (ZR2013HQ008) and the young scientist supportive program of Shandong University of Technology. The funders had no role in the study design, data collection or analysis, the decision to publish or the preparation of the manuscript.

- 1 Dick GW, Kitchen SF, Haddock AJ. Zika virus. I. Isolations and serological specificity. *Trans R Soc Trop Med Hyg* 1952; **46**: 509–520.
- 2 Anderson KB, Thomas SJ, Endy TP. The emergence of Zika virus: a narrative review. *Ann Intern Med* 2016; **165**: 175–183.
- 3 Duffy MR, Chen TH, Hancock WT *et al*. Zika virus outbreak on Yap Island, Federated States of Micronesia. *N Engl J Med* 2009; **360**: 2536–2543.
- 4 Mlakar J, Korva M, Tul N *et al*. Zika virus associated with microcephaly. *N Engl J Med* 2016; **374**: 951–958.
- 5 Broutet N, Krauer F, Riesen M *et al*. Zika virus as a cause of neurologic disorders. *N Engl J Med* 2016; **374**: 1506–1509.
- 6 Carteaux G, Maquart M, Bedet A *et al*. Zika virus associated with meningoencephalitis. *N Engl J Med* 2016; **374**: 1595–1596.
- 7 World Health Organization. *Zika virus and potential complications*. WHO: Geneva, 2016. Available at <http://www.who.int/emergencies/zika-virus/en/>.
- 8 Faria NR, Azevedo Rdo S, Kraemer MU *et al*. Zika virus in the Americas: early epidemiological and genetic findings. *Science* 2016; **352**: 345–349.
- 9 Wang L, Valderramos SG, Wu A *et al*. From mosquitos to humans: genetic evolution of Zika virus. *Cell Host Microbe* 2016; **19**: 561–565.
- 10 Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007; **7**: 214.
- 11 Pan XL, Liu H, Wang HY *et al*. Emergence of genotype I of Japanese encephalitis virus as the dominant genotype in Asia. *J Virol* 2011; **85**: 9847–9853.
- 12 May FJ, Davis CT, Tesh RB *et al*. Phylogeography of West Nile virus: from the cradle of evolution in Africa to Eurasia, Australia, and the Americas. *J Virol* 2011; **85**: 2964–2974.
- 13 Bryant JE, Holmes EC, Barrett AD. Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. *PLoS Pathog* 2007; **3**: e75.
- 14 Muniraju M, Munir M, Parthiban AR *et al*. Molecular evolution of peste des petits ruminants virus. *Emerg Infect Dis* 2014; **20**: 2023–2033.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016