# The Diversification of Zika Virus: Are There Two Distinct Lineages?

Zhen Gong, Xiaoyu Xu, and Guan-Zhu Han*

Jiangsu Key Laboratory for Microbes and Functional Genomics, Jiangsu Engineering and Technology Research Center for Microbiology, College of Life Sciences, Nanjing Normal University, Jiangsu, China

*Corresponding author: E-mail: guanzhu@njnu.edu.cn.

## Abstract

Zika virus (ZIKV) has caused explosive epidemics in the Pacific and the Americas, posing a serious threat to public health. Conventional opinion advocates that ZIKV evolved into two distinct lineages, namely, African and Asian. Descendants of this latter lineage dispersed globally causing major epidemics. However, based on shared amino acid replacements and phylogenetic analyses, it was recently contentiously proposed that the Asian lineage was a direct descendant of the African lineage. To address this contentious issue, we reconstructed a phylogenetic tree of ZIKV using the method based on shared amino acid replacements and found that ZIKV evolved into two distinct lineages. This supports the conventional phylogenetic divergence pattern of ZIKV. Evidence of recombination and sequencing errors was identified among the large collection of ZIKV. As such problematic sequences could confound the phylogenetic analyses, they were removed. Bayesian phylogenetic analyses using the improved sequence data enabled estimates for the divergence time in the past of the African and Asian lineages of ~180 years ago. Moreover, we found that the Asian lineage viruses did not evolve at an elevated rate. Our findings provide additional support for the conventional opinion that the Asian lineage of ZIKV diverged from the African lineage.

**Key words:** phylogenetic analysis, Zika virus, parallel evolution.

Zika virus (ZIKV), a mosquito-borne flavivirus, was first discovered in the serum of a rhesus monkey caged in the Zika Forest of Uganda in 1947 (Dick et al. 1952). Since then, ZIKV was reported sporadically in both Africa and Asia (Haddow et al. 2012; Shen et al. 2016). Recently, ZIKV has undergone a rapid geographic expansion. In 2007, a ZIKV outbreak was reported in Yap Island, Federated States of Micronesia (Duffy et al. 2009). In 2013–2014, French Polynesia in the Southern Pacific experienced a ZIKV outbreak (Cao-Lormeau et al. 2014, 2016). Since the report of autochthonous transmission of ZIKV in Brazil in May 2015 (Kindhauser et al. 2016), ZIKV has spread rapidly throughout the Americas, among which Brazil has the highest number of reported cases of ZIKV. Although ZIKV infection usually leads to no symptom or mild illness (Goodman et al. 2016), ZIKV can also cause a range of neurological disorders, including Guillain–Barré syndrome and microcephaly (Petersen et al. 2016; Krauer et al. 2017). But the risk of microcephaly due to ZIKV infection in pregnancy is relatively low (Cauchemez et al. 2016).

## New Evolutionary Tree of ZIKV?

Numerous conventional phylogenetic analyses of ZIKV genomes reveal the presence of two main viral lineages, that is, African and Asian lineages (Haddow et al. 2012; Faria et al. 2016; Liu et al. 2016; Pettersson et al. 2016; Shen et al. 2016; Wang et al. 2016; Zhu et al. 2016) (fig. 1A and supplementary fig. S1, Supplementary Material online). It should be noted that phylogenetic analyses using E and NS5 genes reveal three major lineages of ZIKV; an additional lineage has been circulating in Africa (designated African II lineage) (Gong et al. 2016; Shen et al. 2016). Unfortunately, no genome sequence has been available for strains of African II lineages to date. Therefore, we still follow the traditional two-lineage nomenclature, Asian and African (actually African I in the new nomenclature) lineages (Gong et al. 2016). However, Yokoyama and Starmer (2017) inferred a rooted phylogenetic tree using a method based on shared amino acid replacements but without any technical detail (fig. 1B). They suggested that MR766, the strain isolated in Uganda in 1947, is basal to all the other ZIKV strains and the Asian and African lineages cannot be readily distinguished. Statistical tests reveal that the topologies
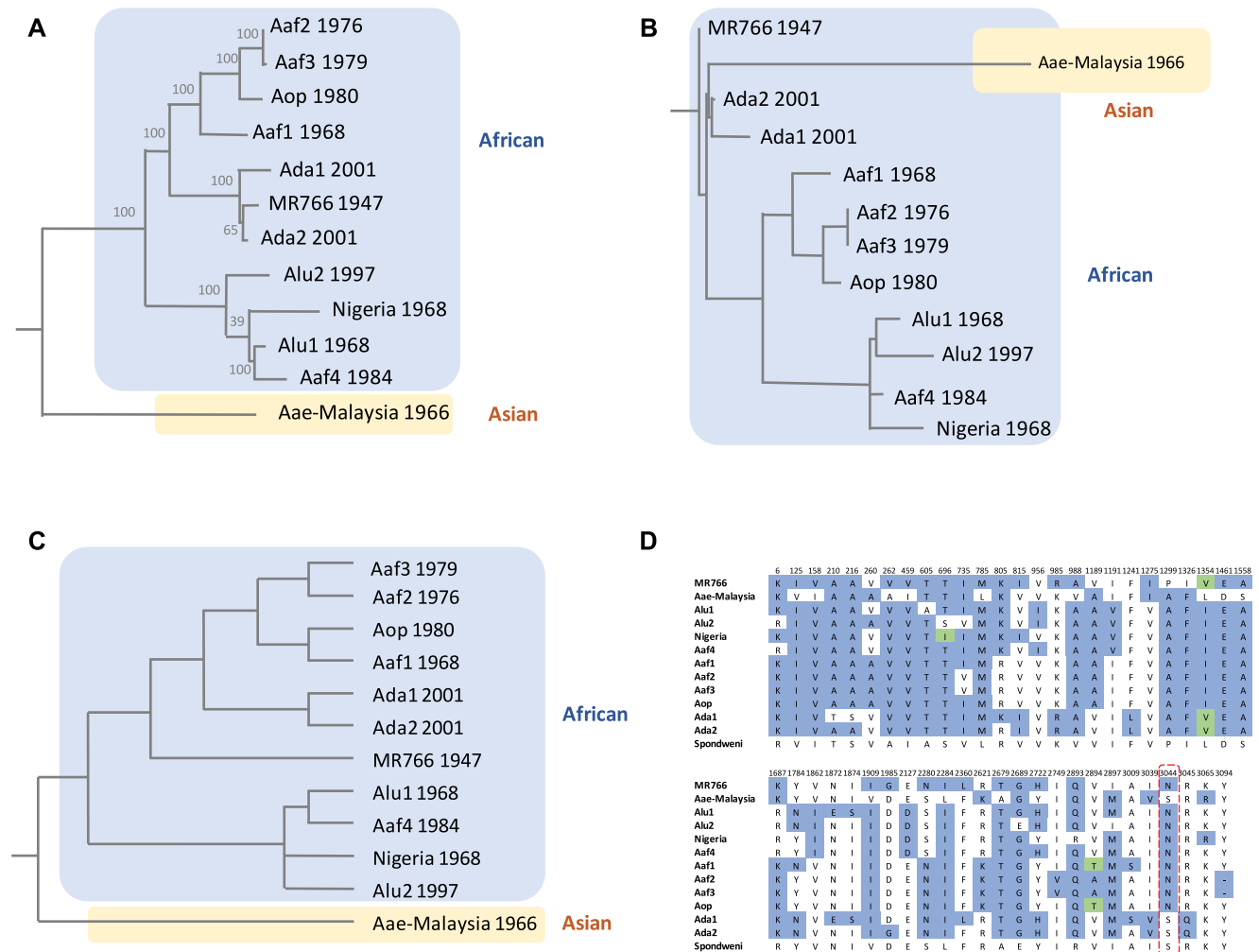
Fig. 1.—Different classifications of ZIKV. The phylogenetic tree (A) was inferred using maximum likelihood method. The conventional Asian and African lineages are indicated by yellow and blue colors. The numbers on the branches represent the bootstrap values. The tree proposed by Yokoyama and Starmer is shown in the right panel (B). The Asian lineage is indicated by yellow color. The topology of the tree (C) was inferred based on the shared amino acid replacements (D). The conventional Asian and African lineages are indicated by yellow and blue colors. For the shared amino acid replacements, the mutations were labeled in colored boxes. The numbers indicate the residue positions in the polyprotein of ZIKV. Strain information (country, year, accession No.): Ada1 (Senegal, 2001, KF383118), Ada2 (Senegal, 2001, KF383119), Aaf1 (African Republic, 1968, KF383115), Aaf2 (Central African Republic, 1976, KF268950), Aaf3 (Central African Republic, 1979, KF268948), Aaf4 (Senegal, 1984, HQ234501), Aae-Malaysia (Malaysia, 1966, KX377336), Alu1 (Senegal, 1968, KF383116), Alu2 (Senegal, 1997, KF383117), Aop (African Republic, 1980, KF268949), MR766 (Uganda, 1947, LC_002520), and Nigeria (Nigeria, 1968, HQ234500).

between the conventional tree and the Yokoyama and Starmer's (YS) tree are significantly different ($P < 0.05$ for the Shimodaira–Hasegawa [SH] test and $P < 0.05$ for the AU test).

The YS tree was reconstructed "by establishing the tree topology based on the shared amino acid replacements in their polyproteins" (Yokoyama and Starmer 2017) and by first "evaluating the number of shared amino acid changes for each pair of sequences" (personal communication with Prof. Yokoyama). Yokoyama and Starmer (2017) proposed that "the ancestral Asian strain diverged from the common ancestor of Ada1 and Ada2 because of one shared amino acid replacement at position 3,040 in the polyprotein" (Yokoyama and Starmer 2017). On close inspection of the

data (supplementary figure 2 in Yokoyama and Starmer 2017), no mutation was indicated in MR766 and thus they used MR766 as the ancestral strain, essentially the outgroup. However, either of these assumptions could be invalid, because: 1) One shared replacement might be caused by convergent evolution (see Chen et al. 2016 for cases in influenza virus) or reversion (Han 2017). Actually, the residue 3,040 (residue 3,044 in fig. 1D) in the Asian strain Aae-Malaysia, Ada1, and Ada2 is likely to represent the ancestral state, rather than a shared amino acid replacement as identified by Yokoyama and Starmer (2017), because the residue in these strains shares the same amino acid as Spondweni virus; 2) The Asian strain shared more replacements with other

African strains than with Ada1 and Ada2 (supplementary figure 2 in Yokoyama and Starmer 2017; fig. 1D); 3) Relative date of sampling is not necessarily an indication of chronological order of appearance, for example, sequences of the 1918 Spanish flu cluster within "modern" flu viruses (Worobey et al. 2014). Therefore, MR766 could not be directly used as the outgroup.

We re-examined the phylogenetic relationships among African and Asian ZIKV strains following the principle of shared amino acid replacements (see Materials and Methods for details). Our phylogenetic tree (fig. 1C) had a very similar topology to the previously published conventional phylogenetic trees and was therefore significantly different from the one proposed by Yokoyama and Starmer (2017). The phylogenetic positions of Ada1, Ada2, and MR766 differ from previous analyses (Shen et al. 2016; Yokoyama and Starmer 2017), possibly because the shared replacements are limited in number and might not be phylogenetically informative. Moreover, we also performed a Bayes Factor (BF) comparison by estimating the ratio of the marginal likelihood of MR766, Ada1, and Ada2 forming a monophyletic group (the conventional tree, fig. 1A; mean marginal likelihood [ln] = −36,628.07) to the marginal likelihood of MR766, Ada1, and Ada2 forming a paraphyletic group (the YS tree, fig. 1B; mean marginal likelihood [ln] = −36,639.33), using the Stepping Stone sampling method. Apparently, the BF > 100 suggests that the conventional tree is more strongly supported by the data. Therefore, these findings provide no evidence for the Asian lineage being a continuation of the African lineage but do support the conventional concept of an African ancestral strain of ZIKV diverging to produce distinct African and Asian lineages.

## Evolutionary Stability or Experimental Errors?

Yokoyama and Starmer (2017) observed that MR766 does not exhibit "a much shorter branch length than many other African strains" in the conventional phylogenetic analysis (fig. 1A) and thus disputed the power of conventional phylogenetic methods. However, numerous previous phylogenetic analyses produced the short branches of sequences from archived samples (e.g., a case of HIV of 1960 in Worobey et al. 2008 and a case of Yersinia pestis from victims of the Black Death in Bos et al. 2011). We believe the absence of short length branches might be explained by reasons other than the choice of phylogenetic methods. Strains MR776 and Ada1, sampled 54 years apart, cluster together with very short length, which seemingly implies a very low evolutionary rate of virus in this cluster. However, similar "evolutionary stasis" was observed in influenza virus and is most likely to be caused by laboratory contamination or other experimental errors (Worobey 2008).

To identify problematic sequences, we performed a comprehensive phylogenetic analysis of 519 ZIKV genomes.

We looked for closely related sequences with noticeably longer or shorter branch lengths than might be expected, arguing that they could have arisen through experimental errors (Worobey 2008). We identified ten MR766 strains that clustered together. However, these strains showed significant genetic variation which could have arisen due to their high passage level under laboratory conditions. For example, strains HQ234498, KU955594, and KU963573 had been passaged >150 times. Moreover, MR766 strain DQ859059 did not cluster with the other MR766 strains. Instead, it clustered with a strain isolated in Uganda in 1962. Accordingly, these problematic sequences were not included in our analyses.

Recombination also confounds phylogenetic analyses, because different regions have different evolutionary histories (Han and Worobey 2011). Therefore, we performed recombination analyses within ZIKV genomes. We identified a total of nine sequences of hybrid origin, among which eight were supported by at least three recombination detection statistics (Table 1). It is unclear whether these sequences represent authentic recombinants, because recombinants can also arise through experimental or sequencing errors (Han and Worobey 2011). Indeed, the observations that potential parental strain for some recombinants is MR766 and the isolate dates between recombinants and their potential parents are tens of years apart make experimental errors a plausible explanation. Nevertheless, these sequences of possible hybrid origin should be excluded from phylogenetic analyses.

## Elevated Evolutionary Rate of Asian Lineage?

To further explore the evolutionary time scale and rate of ZIKV, we performed Bayesian phylogenetic analyses using a clean data set, which excludes the potential problematic and recombinant sequences. The Bayesian phylogenetic analyses support the conventional classification of ZIKV. The median time of the most recent common ancestor (TMRCA) of ZIKV was estimated to be 1,837.8 (95% highest posterior density [95% HPD]: 1,776.5–1,887.7) (fig. 2A), much earlier than 1930–1945 estimated by Yokoyama and Starmer (2017). Previous studies suggested that the Asian lineage seems to have a higher evolutionary rate (Liu et al. 2016; Yokoyama and Starmer 2017). However, we did not find significant difference in the median evolutionary rates among the whole ZIKV population (6.1 [95% HPD: 4.5–7.9] × $10^{-4}$ substitutions per site per year) and the Asian lineage (7.6 [95% HPD: 5.7–9.6] × $10^{-4}$ substitutions per site per year) as claimed by Yokoyama and Starmer (2017) and several other studies (Liu et al. 2016; Pettersson et al. 2016) (fig. 2B). Unfortunately, due to the limited number of African strains, we could not directly estimate the evolutionary rate. Our findings provide no evidence for an elevated rate for Asian lineages. The difference in evolutionary rates observed in previous studies might be caused by the inclusion of more pandemic
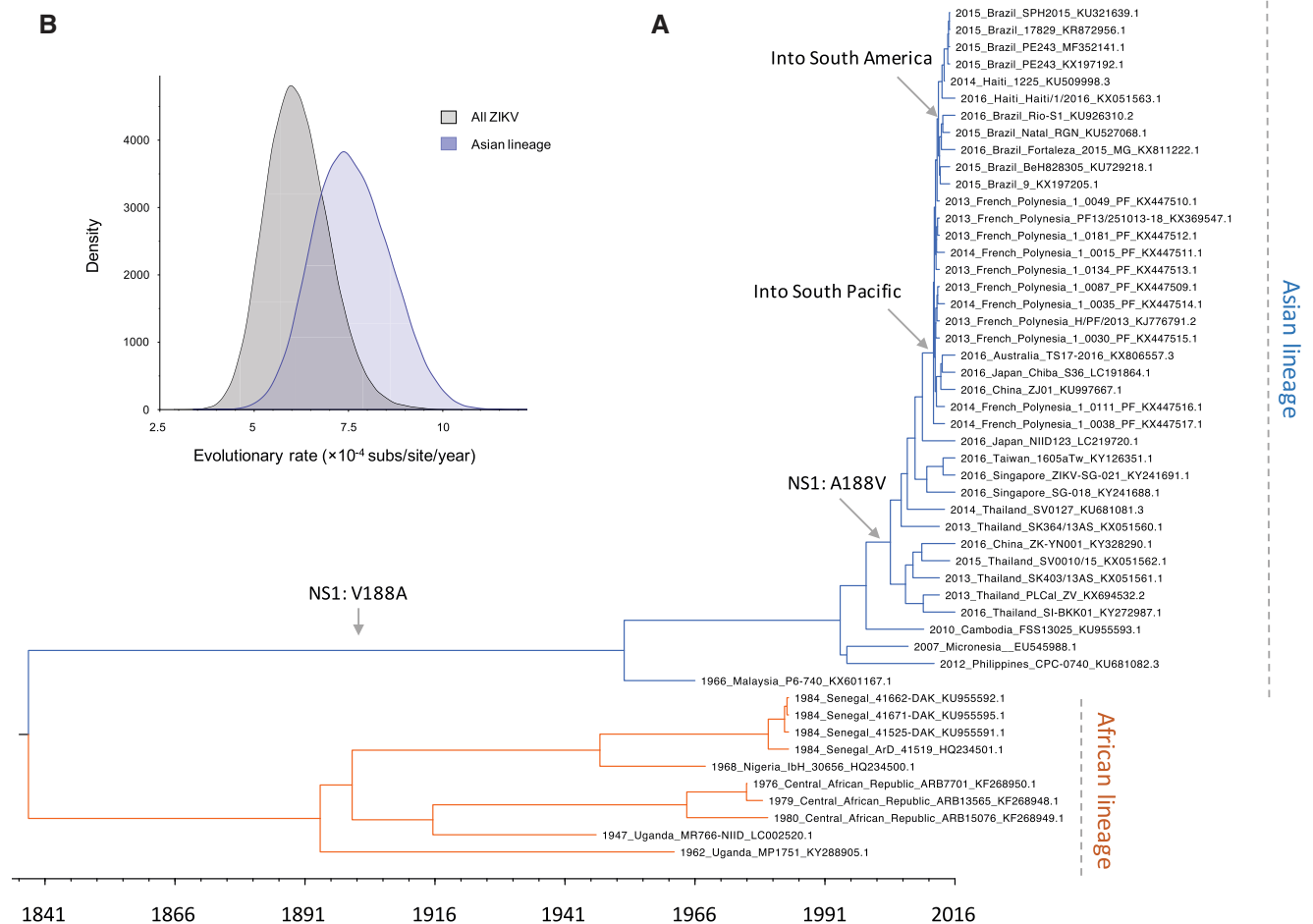
**Fig. 2.**—Evolutionary timescales and rates of ZIKV. (*A*) MCC tree of ZIKV. The tips correspond to the isolate years. The Asian and African lineages are highlighted in blue and orange, respectively. (*B*) The 95% HPD for the distribution of evolutionary rates of ZIKV. The distribution of all ZIKV and Asian lineage is highlighted in gray and blue, respectively.

ZIKV strains. Indeed, the higher evolutionary rate within outbreaks is not unexpected, given sequences within outbreaks were sampled in a relatively short time frame and slightly deleterious mutations were not removed by purifying selection (Duchene et al. 2014; Aiewsakun and Katzourakis 2016; Holmes et al. 2016).

Our time-scaled tree also provides novel insights into the evolutionary history of functionally important amino acids. An amino acid substitution (A188V in nonstructural protein 1 [NS1] protein) is associated with increased infectivity of the ZIKV strain responsible for the epidemic in the Americas in the mosquito vector *Aedes aegypti* (Liu et al. 2017). The A188V substitution was thought to contribute to the recent emergence of ZIKV in the South Pacific and South America (Liu et al. 2017). However, our Bayesian phylogenetic analysis shows that the 188 V arose in Southeast Asia in between 1,998.9 (95%HPD: 1,994.1–2,003.0) and 2,003.6 (95%HPD: 2,000.1–2,006.6), which suggests the substitution occurred earlier than expected. The residue 188 of NS1

protein in strains isolated in South Asia after 2012 are all Valine, suggesting the A188V might have been fixed in the viral population of South Asia. Given the strain with 188 A was isolated in Philippines in 2012, the virus variants with 188 V and 188 A might have been cocirculating for some time. Although it remains unclear what drove the fixation of A188V, it appears to be not directly related to the emergence of the recent ZIKV epidemic, as proposed recently (Han 2017).

## Materials and Methods

### Phylogenetic Analyses Based on Conventional Methods and Shared Amino Acid Replacements

Genome sequences of ZIKV and Spondweni virus were retrieved from GenBank. The accession numbers and isolation information of those viruses are in supplementary figure S2, Supplementary Material online. The nucleotide sequences were aligned using MAFFT (Katoh and Standley 2013) and

**Table 1**

Recombination Analysis of ZIKV Genomes

| Recombinant | Event | Major Parent | Minor Parent | R | G | B | M | C | S | T |
|---|---|---|---|---|---|---|---|---|---|---|
| 1968_Senegal_ArD7117 | 1 | 1984_Senegal_41525-DAK | 1968_Central African_ArB1362 | + | + | + | + | + | − | + |
| 1997_Senegal_ArD128000 | 1 | 1968_Central African_ArB1362 | 1968_Senegal_ArD7117 | + | + | + | + | − | − | + |
| | 2 | 1968_Senegal_ArD7117 | 2001_Senegal_ArD158095 | + | + | + | + | + | + | − |
| | 3 | 1984_Senegal_41525-DAK | 2001_Senegal_ArD158095 | + | + | + | + | − | − | − |
| 2000_Senegal_ArD142623 | 1 | 2001_Senegal_ArD157995 | Unknown | − | − | − | − | − | − | + |
| | 2 | Unknown | 1984_Senegal_ArD41519 | + | + | + | + | − | − | − |
| | 3 | Unknown | 1984_Senegal_ArD41519 | + | + | + | + | + | + | − |
| | 4 | 1984_Senegal_ArD41519 | 2001_Senegal_ArD158095 | + | + | + | − | − | − | + |
| | 5 | 1947_Uganda_MR766 | Unknown | − | + | − | + | − | + | − |
| | 6 | Unknown | 1997_Senegal_ArD128000 | − | − | − | + | + | − | + |
| | 7 | 1997_Senegal_ArD128000 | Unknown | − | − | − | + | − | + | − |
| 2001_Senegal_ArD157995 (Ada2) | 1 | 1947_Uganda_MR766 | Unknown | + | + | + | + | − | − | + |
| | 2 | 1947_Uganda_MR766 | 2000_Senegal_ArD142623 | + | + | + | + | + | + | + |
| | 3 | 2001_Senegal_ArD158095 | 1997_Senegal_ArD128000 | + | + | + | + | + | + | − |
| 2001_Senegal_ArD158095 | 1 | 2001_Senegal_ArD157995 | 1947_Uganda_MR766 | + | − | − | + | + | + | + |
| 2016_Singapore_SG-010 | 1 | 2016_Singapore_SG-030 | Unknown | − | − | − | − | − | − | + |
| 2016_Singapore_SG-030 | 1 | 2016_Singapore_SG-001 | Unknown | + | + | + | + | + | + | + |
| 2016_Singapore_SG-042 | 1 | 2016_Singapore_SG-030 | Unknown | − | − | + | + | − | + | + |
| 2016_Singapore_SG-047 | 1 | 2016_China_ZK-YN001 | 1947_Uganda_MR766 | + | + | + | − | − | + | − |

NOTE.—Recombination detection method: R, RDP; G, GENECONV; B, Bootscan; M, MaxChi; C, Chimaera; S, SiScan; T, 3Seq.

manually edited. The conventional phylogenetic analyses were performed based on both nucleotide and protein sequences using maximum likelihood method with GTR + I + Γ substitution model for DNA and JTT + I + Γ substitution model for protein implemented in MEGA7 (Kumar et al. 2016). The substitution model was selected using jModelTest 2 (Darriba et al. 2012). Support for the tree was assessed using 1,000 bootstrap replicates. The evolutionary tree of ZIKV was also inferred using a method based on shared amino acid replacements. Polyprotein of Spondweni virus was used as the outgroup to identify the shared amino acid replacements in ZIKV. Only the residues where Spondweni virus and some ZIKV strains are the same were considered, because in this case the residue of Spondweni virus is likely to reflect the ancestral state. For each pair of ZIKV, the number of shared amino acid replacements was counted. The tree was inferred by finding the sequences with maximum shared replacements and clustering together sequentially.

## Statistical Tests of Tree Topologies

We compared whether the two competing tree topologies (our conventional tree vs. the YS tree) are significantly different using both the SH test and the AU test (Shimodaira 2002) implemented by PAUP* 4.0a. Moreover, to explore which topology is a better fit to the data, we estimated BF, the ratio of the marginal likelihood of one topology to the marginal likelihood of the competing topology, using the Stepping Stone sampling method (Morey et al. 2016) implemented in

MrBayes 3.2.6 (Ronquist 2012). The GTR + I + Γ substitution model was used. A total of ten million generations were run, with diagnostics frequency set to once every 5,000 generations.

## Large-Scale Phylogenetic and Recombination Analyses

All currently available ZIKV genome sequences (a total of 519) were retrieved from GenBank. The nucleotide sequences were aligned using MAFFT (Katoh and Standley 2013) and manually edited. Due to high computational cost, we reconstructed a comprehensive phylogenetic analysis of all ZIKV genomes using an approximately maximum likelihood method implemented in FastTree 2 (Price et al. 2010) with GTR and CAT model. Recombination events within ZIKV genomes were scanned using 3SEQ method implemented in RDP 4 (Martin et al. 2015). The recombination events were confirmed by other recombination detection methods (RDP, GENECONV, Bootscan, MaxChi, Chimaera, and SiScan) and only the events detected by more than three methods were considered to be positive ones.

## Bayesian Phylogenetic Inference

Bayesian phylogenetic analyses of ZIKV and Asian lineage were reconstructed using BEASTv.1.8.2 (Drummond et al. 2012). Following Liu et al. (2016), we accommodated variation in evolutionary rate among lineages using a lognormal distribution in an uncorrelated relaxed molecular clock model and modelled changes in effective population size through

time using the Bayesian Skyline model (Drummond et al. 2005). The GTR + I + Γ substitution model was used. MCMC chains were run for 200 million generations, sampling every 1,000 generations. The convergence was indicated by effective Sample Size (ESS) > 200. The parameters (median and 95% HPD intervals) were summarized after discarding a 10% burn-in using Tracer. The TreeAnnotator was used to infer maximum clade credibility (MCC) trees.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

## Literature Cited

Aiewsakun P, Katzourakis A. 2016. Time-dependent rate phenomenon in viruses. J Virol. 90(16):7184–7195.

Bos KI, et al. 2011. A draft genome of Yersinia pestis from victims of the Black Death. Nature 478(7370):506–510.

Cao-Lormeau VM, et al. 2014. Zika virus, French Polynesia, South Pacific. Emerg Infect Dis. 20(6):1085–1086.

Cao-Lormeau VM, et al. 2016. Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. Lancet 387(10027):1531–1539.

Cauchemez S, et al. 2016. Association between Zika virus and microcephaly in French Polynesia, 2013-15: a retrospective study. Lancet 387(10033):2125–2132.

Chen H, et al. 2016. Dynamic convergent evolution drives the passage adaptation across 48 years' history of H3N2 influenza evolution. Mol Biol Evol. 33(12):3133–3143.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9(8):772.

Dick GW, Kitchen SF, Haddow AJ. 1952. Zika virus. I. Isolations and serological specificity. Trans R Soc Trop Med Hyg. 46(5):509–520.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 22(5):1185–1192.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 29(8):1969–1973.

Duchene S, Holmes EC, Ho SY. 2014. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. Proc Biol Sci. 281:20140732.

Duffy MR, et al. 2009. Zika virus outbreak on Yap Island, Federated States of Micronesia. N Engl J Med. 360(24):2536–2543.

Faria NR, et al. 2016. Zika virus in the Americas: early epidemiological and genetic findings. Science 352(6283):345–349.

Gong Z, Gao Y, Han GZ. 2016. Zika virus: two or three lineages? Trends Microbiol. 24(7):521–522.

Goodman AB, et al. 2016. Characteristics of children aged <18 years with Zika virus disease acquired postnatally – U.S. States, January 2015-July 2016. MMWR Morb Mortal Wkly Rep. 65:1082–1085.

Haddow AD, et al. 2012. Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. PLoS Negl Trop Dis. 6:e1477.

Han GZ. 2017. A single substitution changes Zika virus infectivity in mosquitoes. Trends Microbiol. 25(8):603–605.

Han GZ, Worobey M. 2011. Homologous recombination in negative sense RNA viruses. Viruses 3(8):1358–1373.

Holmes EC, Dudas G, Rambaut A, Andersen KG. 2016. The evolution of Ebola virus: insights from the 2013-2016 epidemic. Nature 538(7624):193–200.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Kindhauser MK, Allen T, Frank V, Santhana RS, Dye C. 2016. Zika: the origin and spread of a mosquito-borne virus. Bull World Health Organ. 94(9):675–686.

Krauer F, et al. 2017. Zika virus infection as a cause of congenital brain abnormalities and Guillain–Barré syndrome: systematic review. PLoS Med. 14:e1002203.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 33(7):1870–1874.

Liu H, et al. 2016. From discovery to outbreak: the genetic evolution of the emerging Zika virus. Emerg Microbes Infect. 5(10):e111.

Liu Y, et al. 2017. Evolutionary enhancement of Zika virus infectivity in Aedes aegypti mosquitoes. Nature 545(7655):482–486.

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 1(1):vev003.

Morey RD, Romeijn JW, Rouder JN. 2016. The philosophy of Bayes factors and the quantification of statistical evidence. J Math Psychol. 72:6–18.

Petersen LR, Jamieson DJ, Honein MA. 2016. Zika Virus. N Engl J Med. 375(3):294–295.

Pettersson JH, et al. 2016. How did Zika virus emerge in the Pacific Islands and Latin America? mBio 7(5):e01239-16.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5(3):e9490.

Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61(3):539–542.

Shen S, et al. 2016. Phylogenetic analysis revealed the central roles of two African countries in the evolution and worldwide spread of Zika virus. Virol Sin. 31(2):118–130.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol. 51(3):492–508.

Wang L, et al. 2016. From mosquitos to humans: genetic evolution of Zika virus. Cell Host Microbe 19(5):561–565.

Worobey M. 2008. Phylogenetic evidence against evolutionary stasis and natural abiotic reservoirs of influenza A virus. J Virol. 82(7):3769–3774.

Worobey M, et al. 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature 455(7213):661–664.

Worobey M, Han GZ, Rambaut A. 2014. A synchronized global sweep of the internal genes of modern avian influenza virus. Nature 508(7495):254–257.

Yokoyama S, Starmer WT. 2017. Possible roles of new mutations shared by Asian and American Zika viruses. Mol Biol Evol. 34(3):34525–34534.

Zhu Z, et al. 2016. Comparative genomic analysis of pre-epidemic and epidemic Zika virus strains for virological factors potentially associated with the rapidly expanding epidemic. Emerg Microbes Infect. 5:e22.

**Associate editor**: Takashi Gojobori