Possible Roles of New Mutations Shared by Asian and American Zika Viruses

Shozo Yokoyama[*,1] and William T. Starmer[2]

[1]Department of Biology, Emory University, Atlanta, GA
[2]Department of Biology, Syracuse University, Syracuse, NY

*Corresponding author: E-mail: syokoya@emory.edu.
Associate editor: Takashi Gojobori

Abstract

Originating in Africa, the Zika virus (ZIKV) has spread to Asia, Pacific Islands and now to the Americas and beyond. Since the first isolation in 1947, ZIKV strains have been sampled at various times in the last 69 years, but this history has not been reflected in studying the patterns of mutation accumulation in their genomes. Implementing the viral history, we show that the ZIKV ancestor appeared sometime in 1930–1945 and, at that point, its mutation rate was probably less than 0.2×10^{-3}/nucleotide site/year and subsequently increased significantly in most of its descendants. Sustaining a high mutation rate of 4×10^{-3}/site/year throughout its evolution, the Ancestral Asian strain, which was sampled from a mosquito in Malaysia, accumulated 13 mutations in the 3′-untranslated region of RNA stem-loops prior to 1963, seven of which generate more stable stem-loop structures and are likely to inhibit cellular antiviral activities, including immune and RNA interference (RNAi) pathways. The seven mutations have been maintained in all Asian and American strains and may be responsible for serious medical problems we are facing today and offer testable hypotheses to examine their roles in molecular interactions during brain development.

Key words: Zika viruses, evolutionary tree, evolutionary rates, RNA structures.

Introduction

The Zika virus (ZIKV) is a member of family Flaviviridae, genus *Flavivirus*, and is transmitted among humans primarily by *Aedes* mosquitos. The single-stranded positive-sense RNA virus has a length of approximately 10,800 nucleotides with a 5′- and 3′-UTR. It encodes a single polyprotein that undergoes post-translational processing to yield capsid (C), membrane precursor (prM), envelope (E), and seven nonstructural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) (Chambers et al. 1990; Cunha et al. 2016). The ZIKV (MR766) was first isolated in 1947 from a sentinel monkey (*Macaca mulatta*) in the Zika forest of Uganda (Dick et al. 1952). Although dispersing within the African Continent, some strains also spread to Malaysia and then to Philippine and Micronesia, forming a new Asian lineage (Haddow et al. 2012; Faye et al. 2014). After sporadic human infections in Africa and Asia, ZIKV outbreaks were reported in Micronesia (Duffy et al. 2009) in 2007 and French Polynesia (Musso, Nilles, et al. 2014) in 2013–2014. It is suspected that ZIKV reached Brazil during the Polynesian epidemic (Faria et al. 2016) and has spread rapidly throughout the Americas (Chang et al. 2016; Faria et al. 2016).

About 80% of ZIKV infections in humans are asymptomatic and when they appear the symptoms are usually mild, including rash, fever, red eyes, joint pain, and headache (Duffy et al. 2009; Musso, Nhan, et al. 2014; Malone et al. 2016). This description of ZIKV infection had to be changed drastically in the summer of 2015 when the association between ZIKV and microcephaly, Guillan–Barre syndrome and other serious brain disorders was found in the Americas (Calvet et al. 2016; Mlakar et al. 2016; Schuler-Faccini et al. 2016) and retrospectively in French Polynesia (Cauchemez et al. 2016). Because of these findings and the authentication as a direct causative agent by the US Centers for Disease Control and Prevention (CDC) (Rasmussen et al. 2016), ZIKV has become an emerging infectious public threat (Calvet et al. 2016; Cauchemez et al. 2016; Chang et al. 2016; Malone et al. 2016; Mlakar et al. 2016; Rasmussen et al. 2016; Schuler-Faccini et al. 2016). Indeed, using human stem cells (Garcez et al. 2016; Tang et al. 2016) and mice (Cugola et al. 2016; Li et al. 2016; Miner et al. 2016), ZIKV is shown to target neural stem cells (neurospheres), brain organoids, placenta, fetal brains and to cause microcephaly; ZIKV has also been found in human fetal brains (Mlakar et al. 2016). Curiously, only a portion (1–13%) of mothers infected with ZIKV have babies with microcephaly and other brain dysfunctions (Cauchemez et al. 2016; Johansson et al. 2016).

To understand the molecular mechanisms, by which ZIKV causes these serious medical problems, one crucial step is to identify mutations that induce cytopathicity and pathogenicity. One effective way of predicting these mutations is to study how they have accumulated in different ZIKV strains and their possible functions. Expansions of ZIKV strains to various geographic regions and potentially critical mutations have been studied primarily by constructing phylogenetic

Open Access

Article

Fast Track

trees of ZIKV and other flaviviruses (Duffy et al. 2009; Haddow et al. 2012; Cao-Lormeau and Musso 2014; Faye et al. 2014; Chang et al. 2016; Enfissi et al. 2016; Faria et al. 2016; Shen et al. 2016; Zhu et al. 2016). These evolutionary trees show that 1) the strains isolated from Asia, Pacific Islands and the Americas (collectively called Asian strains) and their ancestral African strains form two evolutionarily distinct groups and 2) the numbers of nucleotide substitutions per site (or branch lengths) of various strains measured from the ZIKV ancestor are very similar. However, at least the latter result is in error because MR766 basically stopped evolving in 1947 and American strains accumulated additional mutations for the next 69 years.

As described below, the currently known evolutionary trees of different ZIKV strains have been constructed by basically assuming that various strains evolved with similar speeds (Faria et al. 2016). When this assumption is removed, drastically different tree topologies of ZIKV strains emerge. These new evolutionary trees allow us to properly evaluate the numbers of nucleotide (or amino acid) substitutions per site per year (or evolutionary rates of mutant substitution) for different strains and also identify potentially critical mutations in the Asian strains under the proper historical context.

The goal of this paper is to find a biologically acceptable evolutionary tree of ZIKV strains and try to identify critical mutations. For this purpose, we consider some representative

ZIKV with complete genome sequences and exclude partially sequenced African strains and many completely sequenced American strains without losing generality. Throughout the paper, the strains collected from humans are denoted by the countries of their origin and those isolated from mosquitos are indicated by abbreviated species names. The roots of these trees were determined using the Spondweni virus (SM-6V-1) as the outgroup (indicated by stars, supplementary table S1, Supplementary Material online).

Results and Discussion

Conventional Phylogenetic Trees of ZIKV Strains

Before finding, a new evolutionary tree of ZIKV strains, those of 9 Asian and 11 African ZIKV strains were constructed first by applying the maximum likelihood (ML), neighbour-joining (NJ), and maximum parsimony (MP) methods directly to their nucleotide sequences using the MEGA6 web server (Tamura et al. 2013). The resulting three evolutionary trees show virtually identical topologies and the NJ and ML trees have similar branch lengths (fig. 1A and Supplementary figs. S1A and B, Supplementary Material online). These ML and NJ trees were constructed using the Jukes and Cantor (JC) model (see Materials and Methods). When the gamma distances with the JC model were used (Nei and Kumar 2000), the tree topologies do not differ significantly. For example, the NJ trees with the two distance measures show only two minor
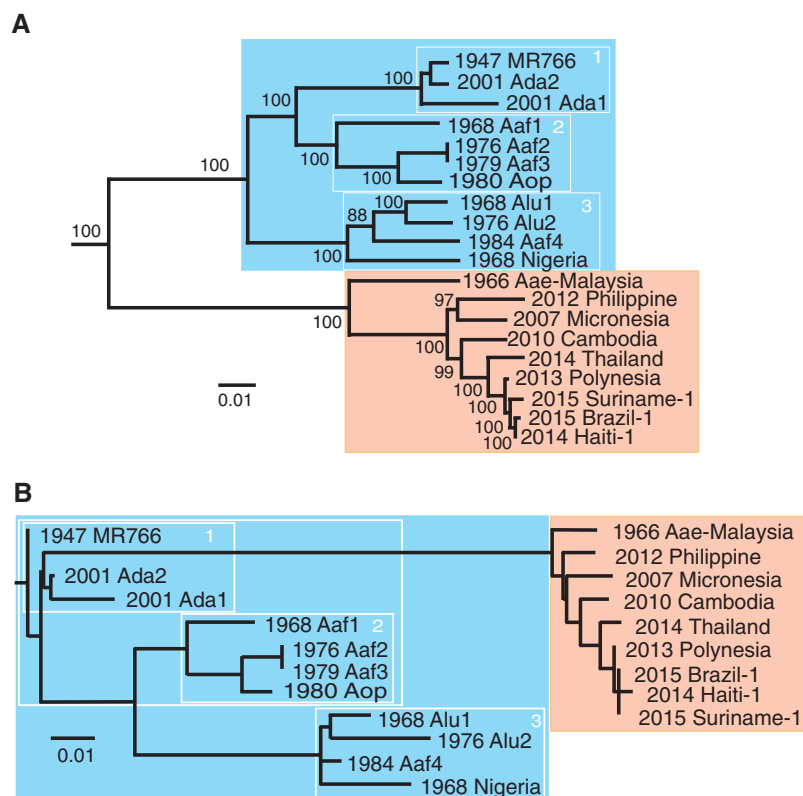

Fig. 1. Phylogenetic trees of 20 representative ZIKV strains. Numbers in front of ZIKV strains indicate the year of isolation. The root was determined using Spondweni virus (SM-6V-1) as the outgroup. (A) The NJ tree, which was obtained by applying the Neighbour-Joining (NJ)-method directly to the sequence data (Tamura et al. 2013). Numbers at various nodes indicate clustering percent supports generated by 1000 bootstrap analyses. Blue and orange rectangles indicate groups 1–3 of African strains and Asian strains which were distinguished using the NJ method, respectively. (B) The PE tree, where the tree topology was determined first and then branch lengths were determined using the ML method (Yang 2007).

differences: MR766 and Ada1 switched their phylogenetic positions in group 1, whereas Alu2 and Aaf4 switched their phylogenetic positions in group 3.

Another phylogenetic tree was obtained using the Bayesian Evolutionary Analyses Sampling Trees (BEAST v1.8.4) method (Drummond and Rambaut 2007). Running 100×10^6 generations with the substitution model (HKY + three codon positions), the clock model (strict) and the tree model (constant), the minimum effective sample size (ESS) of 1795 was obtained (supplementary fig. S1C, Supplementary Material online). Despite assimilating the "time-stamped data," which certainly show a shorter branch length for MR766, the BEAST tree has a very similar topology as those of the ML, MP, and NJ trees (supplementary fig. S1C, Supplementary Material online). The BEAST trees based on the uncorrelated lognormal clock (300×10^6 generations, ESS > 376) and the uncorrelated exponential clock (300×10^6 generations, ESS > 326) are very similar to the BEAST tree with the strict clock model, except that Alu2 and Nigeria are switched within group 3.

All of these trees show that the ancestor of African strains MR766, Ada1, and Ada2 (group 1) split from the common ancestor of Aaf1, Aaf2, Aaf3, and Aop (group 2) and Alu1, Alu2, Aaf4, and Nigeria (group 3). These groupings are depicted in the blue rectangle (fig. 1A). Supported by high bootstrap values (the NJ, ML, and MP trees) and high posterior probabilities (the BEAST tree), these evolutionary steps are, statistically speaking, highly reliable and show that MR766 appeared only recently, which is consistent with the currently known evolutionary trees of ZIKV strains (Haddow et al. 2012; Faye et al. 2014; Enfissi et al. 2016; Faria et al. 2016; Shen et al. 2016; Zhu et al. 2016). In reality, however, MR766 was isolated in 1947 and has a low passage history (Haddow et al. 2012), it is expected to have a much shorter branch length than many other African strains.

To explain these "conventional" tree topologies and similar branch lengths of different strains measured from the ZIKV ancestor, we have to assume that the evolutionary rates of nucleotide substitution (or mutation rates) had to be rigorously adjusted among them so that their branch lengths become very similar at various times of virus isolation—a highly improbable proposition.

A New Evolutionary Tree

To implement the variable sampling history and possibly different evolutionary rates of ZIKV strains, a rooted phylogenetic tree of the same 20 ZIKV strains was constructed first by obtaining its topology based on the shared amino acid replacements (or parallel changes) in their entire polyproteins (supplementary figs. S2 and S3, Supplementary Material online) and then evaluating the numbers of nucleotide substitutions per site for all branches using the ML-method (Yang 2007). In the following, this new evolutionary tree of ZIKV is referred to simply as the parallel evolution (PE) tree (fig. 1B).

The PE tree differs dramatically from the traditional trees (fig. 1A, supplementary fig. S1A–C, Supplementary Material online) and exhibits three major characteristics. First, as expected, MR766 is most distantly related to the other ZIKV strains and consequently the African and Asian strains cannot be clearly distinguished; moreover, MR766 has a very short branch length after it split from the common ancestor of the other ZIKV strains. Second, the ancestors of four African lineages (MR766, Ada1, Ada2, and the rest) and Aae-Malaysia diverged from each other about the same time, strongly suggesting that ZIKV evolution experienced a bottleneck. Third, the branch lengths of different strains, exemplified by Aae-Malaysia and Ada2, which were isolated in 1966 and 2001, respectively, do not reflect the times of their isolation, showing that the evolutionary rates of nucleotide substitution are extremely variable among different strains. However, the PE tree and the conventional trees have one common feature—the American strains can be traced back to the ancestors of Polynesia, Thailand, Cambodia, Micronesia (or Philippine) and Aae-Malaysia, in that order.

In constructing the PE tree, it was assumed that the ancestral Asian strain diverged from the common ancestor of Ada1 and Ada2 because of one shared amino acid replacement at position 3040 in the polyprotein (supplementary fig.S2, Supplementary Material online). Recently, considering the E and NS5 regions separately, it has been claimed that ZIKV strains should be distinguished into two distinct lineages—lineage 1 includes the African strains we are considering (fig. 1A and B), whereas lineage 2 contains ZIKV strains sampled from Senegal and Cote d'Ivoire (Shen et al. 2016). Again, the trees based on the E and NS5 regions show that MR766 appeared only recently and are unrealistic. However, comparing these and the ZIKV sequences considered here (supplementary table S1, Supplementary Material online), we can identify four shared amino acid changes: mutations from I to V (I → V) (at position 3035 in the polyprotein), N → S (at 3040) and R → Q (at 3041) that are shared by Ada1, Ada2, lineage 2 and Asian strains and also Q → H (at 3150) is shared by lineage 2 and African group 3. Hence, the Asian strains seem to have evolved from the common ancestor of Ada1, Ada2, and the lineage 2 strains. To check the effect of this assumption on the final PE tree, another tree was obtained by assuming that the Asian and African ancestors took independent paths after splitting from MR766. This alternate tree (supplementary fig. S4A, Supplementary Material online) is practically indistinguishable from the PE tree (fig. 1B) and the three major characteristics of ZIKV evolution still hold.

The conventional trees, represented by the NJ tree (fig. 1A), can be easily explained from the PE tree (fig. 1B) by considering the highly variable evolutionary rates of ZIKV strains and the methods by which phylogenetic trees are constructed. After all, these trees are constructed basically by searching for strains with smallest sequence differences and clustered together sequentially. Hence, because of their shortest branch lengths measured from the ZIKV ancestor (fig. 1B), MR766, Ada1, and Ada2 appear to belong to the same evolutionary group in the conventional tree (group 1). By the same token, groups 2 and 3 form separate groups and group 1 seem to be more closely related to group 2 than to group 3. Finally, because of the extremely long branch length of Aae-Malaysia, the three groups of African strains and Asian strains seem to form two distinct evolutionary groups in the conventional trees.

Mutation Rates

To evaluate the evolutionary rate of nucleotide substitution of a ZIKV strain, we need to know the duration between its appearance and the time of sampling (or divergence time). Since the branch length of MR766 is virtually zero (4×10^{-6}/site), the PE tree (fig. 1B) suggests that the ZIKV ancestor appeared in 1947, the year of its sampling. However, as described below, MR766 also accumulated some mutations and the PE tree, particularly earlier stages of ZIKV evolution, must be interpreted with caution. In fact, when another tree was constructed considering the numbers of nucleotide substitution per "codon" (Yang 2007), the branch lengths of MR766 and the common ancestor of the rest of ZIKV strains are about the same (1×10^{-4}/codon, supplementary fig. S4B, Supplementary Material online). Hence, it is more likely that the Asian and African strains, MR766 excluded, started to differentiate at around 1947 and the ZIKV ancestor appeared earlier than that (see Origin of ZIKV).

Divergence times (T_1–T_9) and evolutionary rates of nucleotide substitution (a_1–a_{19}) of ZIKV strains were evaluated by assuming that Ada1, Ada2, Aaf1, and Aae-Malaysia diverged from each other in April of 1947 and setting several short branch lengths to zero (fig. 2A). The initial evolutionary steps of the Asian strains were studied by considering Aae-Malaysia and Micronesia together with MR766 and Aaf1 (fig. 2B). Micronesia, rather than Philippine, was used here because their ancestors diverged from Aae-Malaysia at about the same time, but Micronesia was sampled five years earlier than Philippine (supplementary table S1, Supplementary Material online).

Using 10,176 comparable nucleotide sites, the branch length for Aae-Malaysia (L_1) was evaluated to be 0.081/site from $(d_{0,1} - d_{0,12} + d_{1,12})/2$, where $d_{i,j}$s are the branch lengths between strains i and j ($i, j =$ MR766 [strain 0], Aae-Malaysia [strain 1], and Aaf1 [strain 12]) (fig. 2B, supplementary table S1, Supplementary Material online). These nucleotide substitutions occurred during the period between April 947 and July 1966, or 19.25 years, and therefore $a_1 = L_1/19.25 = 4.2 \pm 0.15 \times 10^{-3}$/site/year. Next, the branch lengths between node O and Micronesia (L_3) and between node O and Malaysia ($L_{1,1}$) were estimated to be 0.030/site and 0.013/site, respectively, and $T_1 = 19.25 \times (L_1 - L_{1,1})/L_1 = 16.2$ years and $a_3 = L_3/(60.17 - T_1) = 0.7 \pm 0.04 \times 10^{-3}$/site/year (see Materials and Methods). Similarly, using appropriate formulae (supplementary table S2, Supplementary Material online), T_2–T_9 and the remaining evolutionary rates were evaluated.

The results reveal five characteristics of ZIKV evolution. First of all, the divergence times for the 5 major evolutionary steps of Asian strains vary between 4 and 32 years, whereas those for the 4 major steps of groups 2 and 3 of African strains range between 4 and 12 years (table 1). Hence, it is likely that the ancestor of the currently known American strains split from the ancestors of Polynesia, Thailand, Cambodia, Micronesia (or Philippine), and Aae-Malaysia roughly in 2013, 2007, 1999, 1995, and 1963, respectively. In African group 2 strains, Aop appeared roughly in 1961 and Afa2 and Aaf3 in 1973, whereas Alu1 split from the

ancestor of Aaf1 in 1957 and the other members of group 3 appeared in 1965.

Second, excluding MR766, the 19 strains can be classified into three groups based on their evolutionary rates: 1) Ada2 (0.05×10^{-3}/site/year), 2) Aae-Malaysia, Aaf1, Aaf2, Alu1, and Nigeria (2.1–4.4×10^{-3}/site/year), and 3) the remaining 13 strains (0.2–1.5×10^{-3}/site/year) (table 2, fig. 2A). Hence, the evolutionary rates of ZIKV strains differ by two orders of magnitudes and are far more variable than previously thought (Faria et al. 2016). The extremely variable evolutionary rates among different strains seem to be the major reason why the BEAST tree with the "time-stamped data" results in a typical traditional tree. Third, as expected from the functional constraints of protein-coding genes in DNA and RNA genomes (Gojobori and Yokoyama 1985; Nei 2013), the evolutionary rate tends to be highest at the third position of a codon, followed by the first and second positions of a codon, in that order (table 2). Fourth, in the rapidly changed Aae-Malaysia, mutations occurred uniformly throughout the protein coding-regions; for example, the largest difference between the evolutionary rates of capsid (3.4×10^{-3}/site/year) and propeptide (6.0×10^{-3}/site/year) segments do not differ significantly. However, the evolutionary rates of amino acid replacement for NS2B and NS4A ($\sim 0.4 \times 10^{-3}$/site/year) and capsid (2.7×10^{-3}/site/year) differ significantly ($P < 0.05$) and the evolutionary rate for E (0.8×10^{-3}/site/year) does not differ significantly from the average rate for the polyprotein (1.0×10^{-3}/site/year) (supplementary table S3, Supplementary Material online). These characteristics differ significantly from human immunodeficiency viruses (HIV), in which envelope genes ($\sim 1.0 \times 10^{-2}$/site/year) evolved much faster than the rest of the genome ($\sim 1.0 \times 10^{-3}$/site/year) (Li et al. 1988; Yokoyama 1991). Finally, Brazil-2 suggests that the evolutionary rates of nucleotide substitution at the 5′- and 3′-UTR are statistically indistinguishable from those of the coding region (supplementary table S3, Supplementary Material online).

Origin of ZIKV

To study when the ancestors of various ZIKV strains appeared, the nucleotide sequences of MR766 (strain 0) and Spondweni virus (strain 57) were compared with a designated third strain and the length of the third strain-specific branch (L_{3rd}) was evaluated. For each third strain, the time span between the year of its divergence from MR766 (the time of its appearance) and the year of its isolation in nature was determined by dividing L_{3rd} values by their evolutionary rates (a_{3rd}) and then the year of its actual appearance was estimated by subtracting this L_{3rd}/a_{3rd} value from the year it was sampled. With the exception of Ada1, the branch lengths (L_{3rd}) and the resulting estimates for the origin of ZIKV are not affected much by using either the gamma distance with the JC model or the standard JC model (supplementary table S4, Supplementary Material online).

The results show that Ada2 with the slowest evolutionary rate ($a_{3rd} = 0.05 \times 10^{-3}$/site/year) has a L_{3rd}/a_{3rd} of 56 years, implying that the Ada2 ancestor appeared in 1945 (supplementary table S4, Supplementary Material online). It also
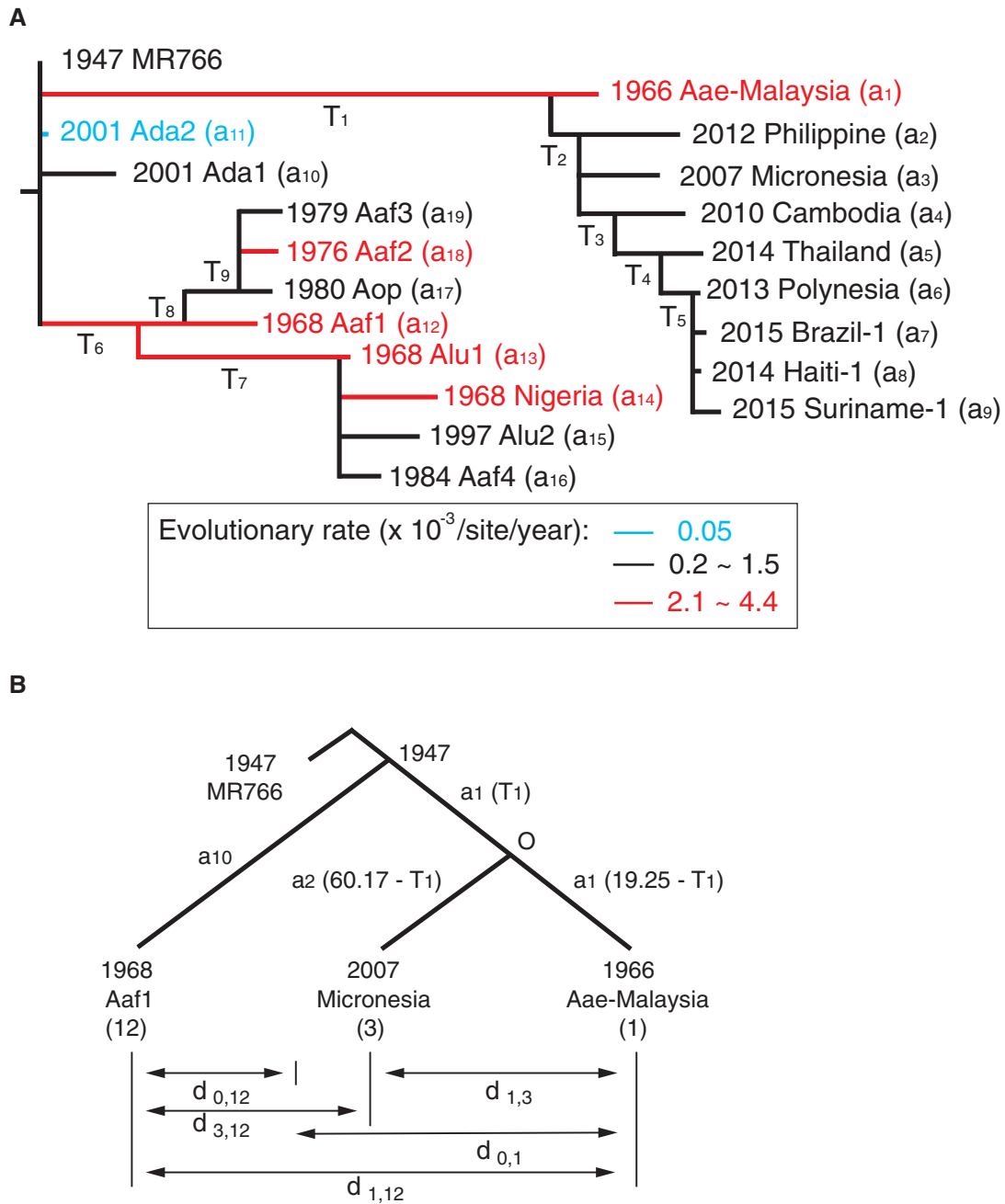
Fig. 2. Estimation of divergence times and evolutionary rates of ZIKV strains. (A) A modified PE tree with divergence times (T_1–T_9) and evolutionary rates of nucleotide substitution per site per year (a_1–a_{19}). Three groups of strains are distinguished by blue, black and red branches with evolutionary rates of 0.05, 0.2–1.5 and 2.0–5.1 × 10^{-3}/site/year, respectively. (B) A phylogenetic relationship among MR766 (strain 0), Aae-Malaysia (strain 1), Micronesia (strain 3), and Aaf1 (strain 12). $d_{i,j}$s are the branch lengths between pairs of strains i and j (i, j = strains 0, 1, 3, and 12). Evolutionary rates a_1, a_3, and a_{12} are followed by divergence times (in parentheses).

suggests that the ancestor of Ada1 appeared in 1928–1941. On the other hand, Nigeria with the fastest evolutionary rate (a_{3rd} = 4.4 × 10^{-3}/site/year) has L_{3rd}/a_{3rd} of 8.8 years and suggests that its ancestor appeared in 1959. Likewise, the evolutionary rates for Aae-Malaysia, Aaf1, and Aop lineages suggest that the ZIKV ancestor appeared sometimes during 1952–1958. Since the ZIKV ancestor appeared prior to the isolation of MR766, the latter results are unrealistic. These results show two things: 1) the mutation rate in ZIKV was initially relatively low, probably less than 0.2 × 10^{-3}/site/

year, but it increased significantly in many descendant lineages and 2) ZIKV appeared roughly in 1930–1945.

Accumulation of New Mutations in Asian Strains
Using the Phylogenetic Analyses by Maximum Likelihood (PAML) with Jones–Taylor–Thornton (JTT), Whelan and Goldman (WAG), and Dayhoff models of amino acid substitution (Yang 2007), amino acid changes at all branches of the PE tree (fig. 1B) were identified. The results consistently show that the ancestors of Aae-Malaysia, Philippine, Cambodia,

Thailand, Polynesia, and 24 American strains have accumulated 33, 14, 3, 5, 2, and 1 new amino acid changes, respectively (supplementary table S5, Supplementary Material online). In this inference, mutants that are polymorphic among Asian strains or those found in the Spondweni virus are excluded because it is possible that the mutant forms are actually ancestral forms. None of these mutations identified here was tested in the previous site-directed mutagenesis experiments using different flaviviruses (Kawano et al. 1993; Amberg and Rice 1999; Mandl et al. 2000; Crill et al. 2007; Zhou et al. 2007; Dong et al. 2010; Winkelmann et al. 2011; Zou et al. 2011; Huang et al. 2014) and some of these mutations could be responsible for causing serious medical problems in human populations. Recent physicochemical analyses suggest that these mutations have a low chance of causing medical problems (Faria et al. 2016). However, we have little knowledge about the individual and nonadditive effects of these mutations of ZIKV function, which need to be determined by mutagenesis experiments.

RNA Structures

Within the 5′-UTR, six mutations occurred in Aae-Malaysia and have been maintained virtually in all Asian strains (supplementary table S6, Supplementary Material online). The RNA-folding structures of the stem-loop (SL-A) of Aae-Malaysia, predicted by *mfold* modeling (Zuker 2003) are similar to that of MR766 (supplementary fig. S5A and B, Supplementary Material online) and no critical mutation can be identified. The 3′-UTR of ZIKV folds into five major SL structures (SL-I, SL-II, SL-III, SL-DB, and 3′-SL) (Zhu et al. 2016). In this region, 16 mutations can be found, 13 of which occurred in Aae-Malaysia and the other three in Philippine have been maintained in all of their descendants (supplementary table S6, Supplementary Material online). The predicted RNA structures of Aae-Malaysia are either identical (SL-II) or very similar (SL-III) to the structure of MR766 (supplementary fig. S5C and D, Supplementary Material online, respectively) and the mutations in these regions do not seem to modify the function of the Asian strains.

Mutations in the rest of the 3′-UTR have distinct possibilities of changing ZIKV functions. The RNA structure of the SL-I of MR766 consists of 24 nucleotides (positions 35 through 58, $\Delta G = -12.9$ Kcal/mol, fig. 3A). Compared with this, Aae-Malaysia with a single mutation (u42c) and Philippine with an additional mutation (u9c) both generate more stable RNA structures with 24-nucleotides ($\Delta G = -15.3$ Kcal/mol) or those with 58-nucleotides ($\Delta G = -14.8$ to -15.6 Kcal/mol) (common segments in blue strips, fig. 3A). Functional importance of this region for flaviviral replication is indicated by chimeric mutants of Dengue type 1 virus (DENV1) (Tajima

Table 1. Divergence Times and the Origins of Various ZIKV Strains.

Divergence Time (in years)	Evolutionary Rate ($\times 10^{-3}$/site/year)	Strain (year of origin)
$T_1 = 16.2$	$a_3 = 0.68$	Philippine and Microneasia (1963)
$T_2 = 32.0$	$a_4 = 0.58$	Cambodia (1995)
$T_3 = 4.3$	$a_5 = 0.67$	Thailand (1999)
$T_4 = 7.9$	$a_6 = 0.53$	Polynesia (2007)
$T_5 = 6.2$	$a_7 = 1.46$	Brazil-1, Haiti-1 and Suriname-1 (2013)
$T_6 = 9.9$	$a_{13} = 3.79$	Alu1 (1957)
$T_7 = 8.0$	$a_{14} = 4.39$	Nigeria, Alu2 and Aaf4 (1965)
$T_8 = 3.7$	$a_{17} = 0.92$	Aop (1951)
$T_9 = 11.6$	$a_{18} = 2.18$	Aaf2 and Aaf3 (1962)

10,176 comparable nucleotide sites were considered.

Table 2. Evolutionary Rates of Nucleotide and Amino Acid Substitutions ($\times 10^{-3}$/site/year).

ZIKV	Divergence Time (years)	Position Within a Codon				Amino Acid
		First	Second	Third	All	
1. Aae-Malaysia	19.25	1.4 ± 0.14	$0.3 \pm 0.07**$	12.6 ± 0.50	4.2 ± 0.15	1.0 ± 0.12
2. Philippine	16.0	0.3 ± 0.07	$0.1 \pm 0.04**$	1.5 ± 0.17	0.7 ± 0.06	0.2 ± 0.05
3. Micronesia	44.0	0.3 ± 0.04	$0.1 \pm 0.03**$	1.9 ± 0.12	0.7 ± 0.04	0.2 ± 0.04
4. Cambodia	15.1	0.3 ± 0.07	$0.1 \pm 0.03**$	1.5 ± 0.17	0.6 ± 0.06	0.1 ± 0.05
5. Thailand	14.8	0.3 ± 0.08	$0.1 \pm 0.05*(**)$	1.6 ± 0.18	0.7 ± 0.07	0.3 ± 0.07
6. Polynesia	6.2	0.2 ± 0.09	$0.05 \pm 0.05**$	1.4 ± 0.26	0.5 ± 0.09	0.1 ± 0.07
7. Brazil-1	1.3	0.9 ± 0.46	$0(**)$	3.3 ± 0.87	1.5 ± 0.33	1.4 ± 0.56
8. Haiti-1	1.1	0.8 ± 0.47	$0(**)$	2.8 ± 0.87	1.3 ± 0.34	1.4 ± 0.60
9. Suriname-1	1.9	0.6 ± 0.32	$0.2 \pm 0.16(**)$	3.5 ± 0.74	1.4 ± 0.27	0.5 ± 0.27
10. Ada1	54	0.2 ± 0.03	$0.2 \pm 0.03(*)$	0.3 ± 0.04	0.2 ± 0.02	0.3 ± 0.04
11. Ada2	54	0.02 ± 0.01	$0.01 \pm 0.01(**)$	0.1 ± 0.02	0.05 ± 0.01	0.01 ± 0.01
12. Aaf1	21	0.6 ± 0.09	$0.2 \pm 0.05**$	5.7 ± 0.30	2.1 ± 0.10	0.4 ± 0.07
13. Alu1	32.1	1.3 ± 0.19	$0.3 \pm 0.08**$	10.6 ± 0.56	3.8 ± 0.19	0.7 ± 0.13
14. Nigeria	3.1	1.6 ± 0.36	$1.3 \pm 0.33(**)$	10.7 ± 0.96	4.4 ± 0.35	2.0 ± 0.40
15. Alu2	32.1	0.4 ± 0.06	$0.1 \pm 0.03**$	0.8 ± 0.09	0.4 ± 0.05	0.2 ± 0.05
16. Aaf4	19.1	0.1 ± 0.04	$0.0 \pm 0.02*(**)$	1.4 ± 0.15	0.5 ± 0.05	0.1 ± 0.03
17. Aop	19.4	0.3 ± 0.07	$0.0 \pm 0.01**$	2.5 ± 0.20	0.9 ± 0.07	0.1 ± 0.04
18. Aaf2	3.8	0.8 ± 0.25	$0.2 \pm 0.11*(**)$	5.7 ± 0.67	2.2 ± 0.24	0.5 ± 0.21
19. Aaf3	6.8	0.5 ± 0.14	$0.1 \pm 0.06**$	3.2 ± 0.37	1.2 ± 0.13	0.3 ± 0.12

*The evolutionary rate at the second position is smaller than that at the first position ($P < 0.05$).
**The evolutionary rate at the second position is smaller than those at the first and third positions ($P < 0.01$).
(*) and (**) The evolutionary rate at the second position is smaller than that at the third position ($P < 0.05$ and $P < 0.01$, respectively).
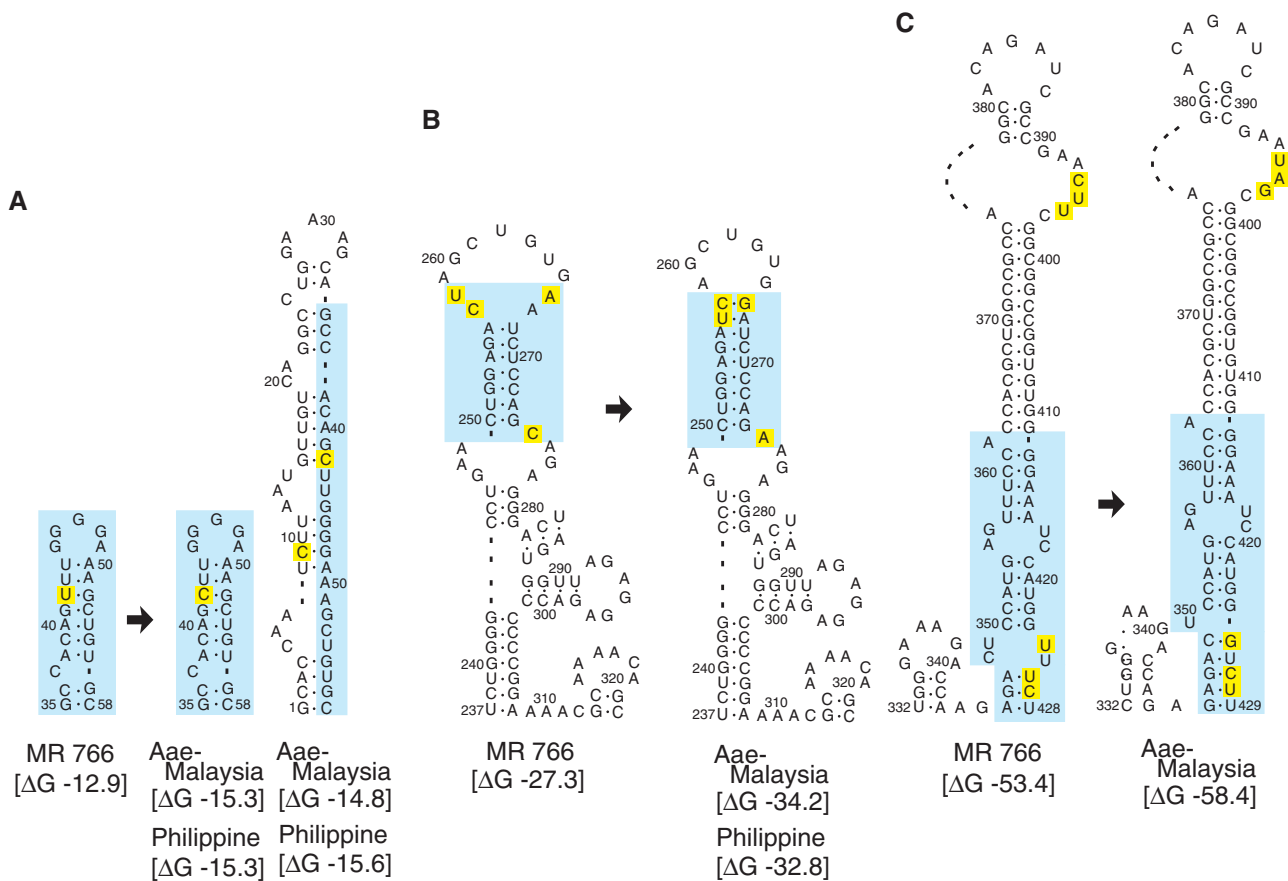
Fig. 3. *mfold* RNA structure models of MR 766 and Asian strains. (*A*) The SL-I structures. (*B*) The SL-DB structures. (*C*) The 3′-SL structures. The comparable segments of the African and Asian strains are shown by blue strips, whereas the mutated sites are highlighted by yellow rectangles. The SL-I, SL-DB, and 3′-SL structures were modeled considering nucleotide sites 1–60, 237–322, and 332–428, respectively, and minor loops for the latter two are not shown.

et al. 2007). Mutations (c257u, u258c, and a266g) make the upper stems of the SL-DB of Aae-Malaysia (and Philippine) longer and more stable ($\Delta G = -34$ to -33 Kcal/mol) than that of MR766 ($\Delta G = -27.3$ Kcal/mol) (blue strips, fig. 3*B*). Similarly, compared with the 3′-SL structure of MR766 ($\Delta G = -53.4$ Kcal/mol), mutations (u424g, u426c, and c427u) of Aae-Malaysia make the bottom stem more stable ($\Delta G = -58.4$ Kcal/mol) (fig. 3*C*). This is the exact segment of DENV2 chimeric mutants that grew well in monkey kidney cells but have restricted growth in cultured mosquito cells (Zeng et al. 1998).

The UTR of flaviviruses are known to be required not only for such basic functions as viability, replication and translation (Pijlman et al. 2008; Roby et al. 2014; Villordo et al. 2016) but also for inhibiting innate immune pathways (Manokaran et al. 2015) and small interference RNA- and micro RNA-induced RNAi pathways (Schnettler et al. 2012; Roby et al. 2014), all of which lead to virus-induced cytopathicity and pathogenicity. The effects of the subgenomic RNA products, which include 3′-SL, at the 3′-UTR of flaviviruses on the viral function and host adaptation have been studied using chimeric mutants that recomposed of the corresponding segments of different strains (Men et al. 1996; Zeng et al. 1998; Alvarez et al. 2005; Tajima et al. 2007; Blaney et al. 2008; Tumban et al. 2011).

These observations strongly suggest that the seven critical mutations (u42c, c257u, u258c and a266g, u424g, u426c, and c427u), which occurred in Aae-Malaysia prior to 1963, may be causing the currently known medical problems. It seems paradoxical that the association between ZIKV infection and microcephaly and other serious brain disorders has not been found prior to the 2013–2014 French Polynesian epidemic (Malone et al. 2016). However, it can be easily imagined that the ZIKV-disease associations were simply missed in earlier days, as we witnessed in the original survey of the French Polynesian epidemic. The geographic regions of ZIKV outbreaks coincide with the areas at risk of DENV and Chikungunya virus (CHIKV) outbreaks (Cao-Lormeau and Musso 2014; Shen et al. 2016). These viruses may cause various medical symptoms by interacting with each other (Fagbami et al. 1988; Gérardin et al. 2014). Before, we can fully study such interactions in disease developments, it seems critical to clarify the role of each virus in causing a specific disease, like microcephaly. Fortunately, some animal models for ZIKV infection have been developed (Cugola et al. 2016; Li et al. 2016; Miner et al. 2016). At this stage of ZIKV research, the seven point mutations in the 3′-UTR, possibly together with some of the mutations that occurred in the polyproteins of Asian strains, offer a set of specific testable hypotheses

(rather than using chimeric mutants) to examine how ZIKV started causing a variety of problems in brain development.

Materials and Methods

Construction of Phylogenetic Trees of ZIKV Strains

The "conventional" rooted evolutionary trees of 20 representative strains were constructed by applying ML, neighbour-joining (NJ), and maximum parsimony (MP) methods directly to their nucleotide sequence data together with the outgroup, Spondweni virus (SM-6V-1), using the MEGA6 web server (Tamura et al. 2013). In constructing the ML and NJ trees, branch lengths (d) were estimated from the proportion of different nucleotides per site (p_1) using the relationship $d = -(3/4)\ln[1 - (4/3)p_1]$ under the JC model (Jukes and Cantor 1969) and the gamma distribution with the JC model; the standard error of d was calculated from $[\{9p_1(1 - p_1)\}/\{(3 - 4p_1)^2 n\}]^{1/2}$, where n is the number of nucleotide sites compared. The number of amino acid replacements per site (K) was estimated from the proportion of different amino acids per site (p_2), $K = \ln(1 - p_2)$, and the standard error was calculated from $[p_2/(1 - p_2)n]^{1/2}$, where n is the number of amino acid compared (Nei and Kumar 2000). The Bayesian Evolutionary Analyses Sampling Trees (BEAST) analyses were also used to incorporate the variable times of virus sampling (or "time-stamped data") (Drummond and Rambaut 2007). For each run, we tried to ensure and achieve the ESS of more than 200.

The new phylogenetic trees were obtained first by establishing the tree topology based on the shared amino acid replacements in their polyproteins (supplementary figs. S2 and S3, Supplementary Material online) and then by evaluating the numbers of nucleotide substitutions per site for all branches applying the ML method (Yang 2007). Using the ML-based Bayesian method (Yang 2007) with JTT, WAG, and Dayhoff models of amino acid substitution, ancestral amino acids at all nodes of the PE tree (fig. 2A) were inferred and amino acid replacements have been identified.

Estimation of Divergence Times and the Evolutionary Rates

The evolutionary rate of nucleotide substitution of each ZIKV strain was estimated by dividing its branch length by the time span that the virus was allowed to evolve. Divergence time estimates were based on the ML tree determined from the numbers of nucleotide substitution per codon (supplementary fig. S4B, Supplementary Material online). It was assumed that the ancestors of Ada1, Ada2, Aaf1, and Malaysia diverged from each other in April 1947. The divergence times (T_1–T_9) and evolutionary rates (a_1–a_{19}) (fig. 2A) were estimated in the following fashion.

For Asian strains, the branch length for Aae-Malaysia (L_1) was first evaluated from $(d_{0,1} - d_{0,12} + d_{1,12})/2$, where $d_{0,1}$, $d_{0,12}$, and $d_{1,12}$ are the branch lengths between MR766 (strain 0) and Aae-Malaysia (strain 1), between MR766 and Aaf1 (strain 12) and between Aae-Malaysia and Adf1, respectively (fig. 2B). Aae-Malaysia appeared in April 1947 and sampled in July 1966, or the divergence time of 19.25 years, and therefore

a_1 is given by $L_1/19.25$ (supplementary table S2, Supplementary Material online). L_1 can also be evaluated using other formulae such as $(d_{1,10} + d_{1,12} - d_{10,12})/2$ and $(d_{1,11} + d_{1,12} - d_{11,12})/2$. However, some values of T_6–T_9 become negative when the L_1 values are estimated using these formulae and thus they were not used.

To evaluate the evolutionary rates of Phillipine, Micronesia, and Cambodia, we need to estimate T_1 (fig. 2A). The branch lengths between node O and Micronesia (L_3) and between node O and Aae-Malaysia ($L_{1,1}$) were estimated from $(d_{1,3} - d_{1,12} + d_{3,12})/2$ and $(d_{1,3} + d_{1,12} - d_{3,12})/2$, respectively (fig. 2B). In this calculation, Micronesia, rather than Philippine or Cambodia, was used because Micronesia was sampled earlier than the other two strains, which minimize the possible effect of random genetic drift on the analyses. Then, using relationships $a_1 (19.25 - T_1) = L_{1,1}$ and $a_3 (60.17 - T_1) = L_3$, we obtain $T_1 = 19.25 (L_1 - L_{1,1})/L_1$ and $a_3 = L_3/(60.17 - T_1)$ (supplementary table S2, Supplementary Material online). Similarly, pairs of T_2 and a_4, T_3 and a_5, T_4 and a_6 and finally T_5 and a_7 were evaluated together. T_2 can be evaluated using the branch length of Philippine or Cambodia; however, the branch leading to Cambodia was used to assure positive values for T_3, T_4, and T_5. Knowing these divergence times, a_2, a_8, and a_9 can also be evaluated (supplementary table S2, Supplementary Material online).

For African strains, a_{10} (Ada1) and a_{11} (Ada2) were estimated from $[(d_{10,11} + d_{10,12} - d_{11,12})/2]/54$ and $[(d_{10,11} - d_{10,12} + d_{11,12})/2]/54$, respectively (supplementary table S2, Supplementary Material online). For African group 2 and 3 strains (fig. 2B), $a_{12} = [(-d_{10,11} + d_{10,12} + d_{11,12})/2]/21$ was first evaluated and then, using relationships $a_{12} (21 - T_6) = L_{12,1}$ and $a_{13} (21 - T_6) = L_{13}$, T_6 and a_{13} were evaluated from $21(L_{13} - L_{12,1})/L_{13}$ and $L_{13}/(21 - T_6)$, respectively, where $L_{13} = (-d_{11,12} + d_{11,13} + d_{12,13})/2$ and $L_{12,1} = (d_{11,12} - d_{11,13} + d_{12,13})/2$. Continuing the process, T_7 and a_{14}, T_8, and a_{17} and finally T_9 and a_{18} were evaluated together and, eventually, a_{15}, a_{16}, and a_{19} were determined. In evaluating all of these evolutionary rates, the branch lengths at all positions of each codon were used so that the effects of sampling biases are minimized.

Inferring Amino Acid and Nucleotide Substitutions and RNA-Folding Modeling

Using the PAML with JTT, WAG, and Dayhoff substitution models (Yang 2007), amino acid replacements at all nodes of the PE tree (fig. 2A) were inferred initially by considering the 21 strains (indicated by *, supplementary table S1, Supplementary Material online) and the maintenance of these mutations in their descendants were studied considering all of the remaining 36 strains. Because of the restriction on the currently available sequence data, the patterns of mutation accumulation in the 5′- and 3′-UTR in Asian strains were constrained to 20 and 13 strains, respectively (indicated by † and ‡, respectively, supplementary table S1, Supplementary Material online).

RNA-folding structures of a stem-loop (SL-A) in the 5′-UTR and those of (SL-I, SL-II, SL-III, SL-DB, and 3′-SL) in the 3′-UTR were evaluated using the *mfold* web server (http://mfold.rna.albany.edu) (Zuker 2003). In modeling SL-I, SL-II, SL-

III, SL-DB, and 3′-SL, nucleotide sites 1–106 for the 5′-UTR and 1–60, 101–170, 180–235, 237–322, and 332–428 are considered, respectively.

Supplementary Material

Acknowledgments

References

Alvarez DE, De Lella Ezcurra AL, Fucito S, Gamarnik AV. 2005. Role of RNA structures present at the 3VUTR of dengue virus on translation, RNA synthesis, and viral replication. *Virology* 339:200–212.

Amberg SM, Rice CM. 1999. Mutagenesis of the NS2B-NS3-mediated cleavage site in the flavivirus capsid protein demonstrates a requirement for coordinated processing. *J Virol.* 73:8083–8094.

Blaney JE, Sathe NS, Goddard L, Hanson CT, Romero TA, Hanley KA, Murphy BR, Whitehead SS. 2008. Dengue virus type 3 vaccine candidates generated by introduction of deletions in the 3′ untranslated region (3′-UTR) or by exchange of the DENV-3 3′-UTR with that of DENV-4. *Vaccine* 26:817–828.

Calvet G, Aguiar RS, Melo ASO, Sampaio SA, de Filippis I, Fabri A, Araujo ESM, de Sequeira PC, de Mendonça MCL, de Oliveira L, et al. 2016. Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. *Lancet Infect Dis.* 16:653–660.

Cao-Lormeau V-M, Musso D. 2014. Emerging arboviruses in the Pacific. *Lancet* 384:1571–1572.

Cauchemez S, Besnard M, Bompard P, Dub T, Guillemette-Artur P, Eyrolle-Guignot D, Salje H, Van Kerkhove MD, Abadie V, Garel C, et al. 2016. Association between Zika virus and microcephaly in French Polynesia, 2013–15: a retrospective study. *Lancet* 387:2125–2132.

Chambers TJ, Hahn CS, Galler R, Rice CM. 1990. Flavivirus genome organization, expression, and replication. *Annu Rev Microbiol.* 44:649–688.

Chang C, Ortiz K, Ansari A, Gershwin ME. 2016. The Zika outbreak of the 21st century. *J Autoimmun.* 68:1–13.

Crill WD, Trainor NB, Chang G-J. 2007. A detailed mutagenesis study of flavivirus crossreactive epitopes using West Nile virus-like particles. *J Gen Virol.* 88:1169–1174.

Cugola FR, Fernandes LR, Russo FB, Freitas BC, Dias JLM, Guimarães KP, Benazzato C, Almeida N, Pignatari GC, Romero S, et al. 2016. The Brazilian Zika virus strain causes birth defects in experimental models. *Nature* 534:267–271.

Cunha MS, Esposito DLA, Rocco IM, Maeda AY, Vasami FGS, Nogueira JS, de Souza RP, Suzuki A, Addas-Carvalho M, Barjas-Castro ML, et al. 2016. First complete genome sequence of Zika virus (Flaviviridae, Flavivirus) from an autochthonous transmission in Brazil. *Genome Announc.* 4:e00032–e00016.

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

Dong K, Liu L, Zou G, Zhao Y, Li Z, Lim SP, Shi P-Y, Li H. 2010. Structural and functional analyses of a conserved hydrophobic pocket of flavivirus methyltransferase. *J Biol Chem.* 285:32586–32595.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7:214.

Duffy MR, Chen T-H, Hancock WT, Powers AM, Kool JK, Lanciotti RS, Pretrick M, Marfel M, Holzbauer S, Dubray C, et al. 2009. Zika virus outbreak on Yap Island, Federated States of Micronesia. *New Engl J Med.* 360:2536–2543.

Enfissi A, Codrington J, Roosblad J, Kazanji M, Rousset D. 2016. Zika virus genome from the Americas. *Lancet* 387:227–228.

Fagbami A, Halstead SB, Marchette N, Larsen K. 1988. Heterologous flavivirus infection-enhancing antibodies in sera of Nigerians. *Am J Trop Med Hyg.* 38:205–207.

Faria NR, da Silva Azevedo RS, Kraemer MUG, Souza R, Cunha MS, Hill SC, Thézé j, Bonsall MB, Bowden TA, Rissanen I, et al. 2016. Zika virus in the Americas: early epidemiological and genetic findings. *Science* 352:345–349.

Faye O, Freire CCM, Iamarino A, Faye O, de Oliveira JVC, Diallo M, Zanotto PMA, Sall AA. 2014. Molecular evolution of Zika virus during Its emergence in the 20th Century. *PLOS Negl Trop Dis.* 8:e2636.

Garcez PP, Loiola EC, da Costa RM, Higa LM, Trindade P, Delvecchio R, Nascimento JM, Brindeiro R, Tanuri R, Rehen SK. 2016. Zika virus impairs growth in human neurospheres and brain organoids. *Science* 352:816–818.

Gérardin P, Sampériz S, Ramful D, Boumahni B, Bintner M, Alessandri JL, Carbonnier M, Tiran-Rajaoefera I, Beullier G, Boya I, et al. 2014. Neurocognitive outcome of children exposed to perinatal mother-to-child Chikungunya virus infection: the CHIMERE cohort study on Reunion Island. *PLOS Negl Trop Dis.* 8:e2996.

Gojobori T, Yokoyama S. 1985. Rates of evolution of the retroviral oncogene of Moloney murine sarcoma virus and of its cellular homologues. *Proc Natl Acad Sci USA.* 82:4198–4201.

Haddow 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.

Huang Y-JS, Nuckols JT, Horne KM, Vanlandingham D, Lobigs M, Higgs S. 2014. Mutagenesis analysis of T380R mutation in the envelope protein of yellow fever virus. *Virology J.* 11:60.

Johansson MA, Mier-y-Teran-Romero L, Reefhuis J, Gilboa S, Hills SL. 2016. Zika and the risk of microcephaly. *New Engl J Med.* 375:1–4.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro HN, editor. Mammalian protein metabolism. New York: Academic Press. p. 21–132.

Kawano H, Rostapshov V, Rosen L, Lai C-J. 1993. Genetic determinants of Dengue type 4 virus neurovirulence for mice. *J Virol.* 67:6567–6575.

Li C, Xu D, Ye Q, Hong S, Jiang Y, Liu X, Zhang N, Shi L, Qin C-F, Xu Z. 2016. Zika virus disrupts neural progenitor development and leads to microcephaly in mice. *Cell Stem Cell* 19:1–7.

Li WH, Tanimura M, Sharp PM. 1988. Rates and dates of divergence between AIDS virus nucleotide sequences. *Mol Biol Evol.* 5:313–330.

Malone RW, Homan J, Michael V, Callahan MV, Glasspool-Malone J, Damodaran L, De Bernardi Schneider A, Zimler R, Talton J, Cobb RR, Ruzic I, et al. 2016. Zika virus: medical countermeasure development challenges. *PLOS Negl Trop Dis.* 10:e0004530.

Mandl CW, Allison SL, Holzmann H, Meixner T, Heinz FX. 2000. Attenuation of tick-horne encephalitis virus by structure-based site-specific mutagenesis of a putative flavivirus receptor binding site. *J Virol.* 74:9601–9609.

Manokaran G, Finol E, Wang C, Gunaratne J, Bahl J, Ong EZ, Tan HC, Sessions OM, Ward AM, Gubler DJ, et al. 2015. Dengue subgenomic RNA binds TRIM25 to inhibit interferon expression for epidemiological fitness. *Science* 350:217–221.

Men R, Bray M, Clark D, Chanock RM, Lai C-J. 1996. Dengue type 4 virus mutants containing deletions in the 3′ noncoding region of the RNA genome: analysis of growth restriction in cell culture and altered viremia pattern and immunogenicity in rhesus monkeys. *J Virol.* 70:3930–3937.

Miner JJ, Cao B, Govero J, Smith AM, Fernandez E, Cabrera OH, Garber C, Noll M, Klein RS, Noguchi KK, et al. 2016. Zika virus infection during pregnancy in mice causes placental damage and fetal demise. *Cell* 165:1081–1091.

Mlakar J, Korva M, Tul N, Popović M, Poljšak-Prijatelj M, Mraz J, Kolenc M, Resman Rus K, Vipotnik TV, Vodušek VF, et al. 2016. Zika virus associated with microcephaly. *New Engl J Med* 374:951–958.

Musso D, Nhan T, Robin E, Roche C, Bierlaire D, Zisou K, Yan AS, Cao-Lormeau V-M, Broult J. 2014a. Potential for Zika virus transmission through blood transfusion demonstrated during an outbreak in French Polynesia, November 2013 to February 2014. *Euro Surveill* 19:pii/20761.

Musso D, Nilles EJ, Cao-Lormeau V-M. 2014b. Rapid spread of emerging Zika virus in the Pacific area. *Clin Microbiol Infect.* 20:O595–O596.

Nei M. 2013. Mutation-driven evolution. Oxford (UK): Oxford University Press.

Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. Oxford (UK): Oxford University Press.

Pijlman GP, Funk A, Kondratieva N, Leung J, Torres S, van der Aa L, Liu WJ, Palmenberg AC, Shi P-Y, Hall RA, Khromykh AA. 2008. A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. *Cell Host Microbe* 4:579–591.

Rasmussen SA, Jamieson DJ, Honein MA, Petersen LR. 2016. Zika virus and birth defects—reviewing the evidence for causality. *New Engl J Med.* 374:1981–1987.

Roby JA, Pijlman GP, Wilusz J, Khromykh AA. 2014. Noncoding subgenomic flavivirus RNA: multiple functions in West Nile virus pathogenesis and modulation of host responses. *Viruses* 6:404–427.

Schnettler E, Sterken MG, Leung JY, Metz SW, Geertsema C, Goldbach RW, Vlak JM, Kohl A, Khromykh AA, Pijlman GP. 2012. Noncoding flavivirus RNA displays RNA interference suppressor activity in insect and mammalian cells. *J Virol.* 86 (24): 13486–13500.

Schuler-Faccini L, Ribeiro EM, Feitosa IML, Horovitz DDG, Cavalcanti DP, Pessoa A, Doriqui MJR, Ivanildo Neri J, Monteiro de Pina Neto J, Wanderley HYC, Force BMGSZET, et al. 2016. Possible Association Between Zika Virus Infection and Microcephaly — Brazil, 2015. *Morb Mortal Wkly Rep.* 65:59–62.

Shen S, Shi J, Wang J, Tang S, Wang H, Hu Z, Deng F. 2016. Phylogenetic analysis revealed the central roles of two African countries in the evolution and worldwide spread of Zika virus. *Virol Sin.* 31:118–130.

Tajima S, Nukui Y, Takasaki T, Kurane I. 2007. Characterization of the variable region in the 3′ non-translated region of dengue type 1 virus. *J Gen Virol.* 88:2214–2222.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Bio. Evol.* 30:2725–2729.

Tang H, Hammack C, Ogden SC, Wen Z, Qian X, Li Y, Yao B, Shin J, Zhang F, Lee EM, et al. 2016. Zika virus infects human cortical neural progenitors and attenuates their growth. *Cell Stem Cell* 18:1–4.

Tumban E, Mitzel DN, Maes NE, Hanson CT, Whitehead SS, Hanley KA. 2011. Replacement of the 39 untranslated variable region of mosquito-borne dengue virus with that of tickborne Langat virus does not alter vector specificity. *J Gen Virol.* 92:841–848.

Villordo SM, Carballeda JM, Filomatori CV, Gamarnik AV. 2016. RNA structure duplications and Flavivirus host adaptation. *Trends Microbiol.* 24:270–283.

Winkelmann ER, Widman DG, Suzuki R, Mason PW. 2011. Analyses of mutations selected by passaging a chimeric flavivirus identify mutations that alter infectivity and reveal an interaction between the structural proteins and the nonstructural glycoprotein NS1. *Virology* 421:96–104.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.

Yokoyama S. 1991. Molecular evolution of human immunodeficiency viruses and related retroviruses. In: Selander RK, Clark AG, Whittam TS, editors. Evolution at the molecular level. Sunderland (MA): Sinauer.

Zeng L, Falgout B, Markoff L. 1998. Identification of specific nucleotide sequences within the conserved 3′-SL in the Dengue type 2 virus genome required for replication. *J Virol.* 72:7510–7522.

Zhou Y, Ray D, Zhao Y, Dong H, Ren S, Li Z, Guo Y, Bernard KA, Shi P-Y, Li H. 2007. Structure and function of flavivirus NS5 methyltransferase. *J Virol.* 81:3891–3903.

Zhu Z, Chan JF-W, Tee K-M, Choi GK-Y, Lau SK-P, Woo PC-Y, Tse H, Yuen K-Y. 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.* e22.

Zou G, Chen Y-L, Dong H, Lim CC, Yap LJ, Yau YH, Shochat SG, Lescar J, Shi P-Y. 2011. Functional analysis of two cavities in flavivirus NS5 polymerase. *J Biol Chem.* 286:14362–14372.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31:3406–3415.