Risk assessment and genomic characterization of Zika virus in China and its surrounding areas

Rong-Fei Liu[1], Zhen-Jian He[1,2], Peng Mei[1], Jia-Cheng Xi[1], Xu-Dong Cao[3], Li-Hong Yuan[4], Jia-Hai Lu[1,2,5]

[1]School of Public Health, Sun Yat-sen University, Guangzhou, Guangdong 510080, China;

[2]Key Laboratory of Tropical Disease Control (Sun Yat-sen University), Ministry of Education, Guangzhou, Guangdong 510080, China;

[3]Department of Chemical and Biological Engineering, University of Ottawa, Ottawa, Ontario, Canada;

[4]School of Life Sciences and Biopharmaceutics, Guangdong Pharmaceutical University, Guangzhou, Guangdong 510006, China;

[5]One Health Center of Excellence for Research & Training, Sun Yat-sen University, Guangzhou, Guangdong 510080, China.

Abstract

Background: Zika virus (ZIKV) has emerged as a global pathogen causing significant public health concerns. China has reported several imported cases where ZIKV were carried by travelers who frequently travel between China and ZIKV-endemic regions. To fully characterize the ZIKV strains isolated from the cases reported in China and assess the risk of ZIKV transmission in China, comprehensive phylogenetic and genetic analyses were performed both on all ZIKV sequences of China and on a group of scientifically selected ZIKV sequences reported in some of the top interested destinations for Chinese travelers.

Methods: ZIKV genomic sequences were retrieved from the National Center for Biotechnology Information database through stratified sampling. Recombination event detection, maximum likelihood (ML) phylogenetic analysis, molecular clock analysis, selection pressure analysis, and amino acid substitution analysis were used to reconstruct the epidemiology and molecular transmission of ZIKV.

Results: The present study investigated 18 ZIKV sequences from China and 70 sequences from 16 selected countries. Recombination events rarely happens in all ZIKV Asian lineage. ZIKV genomes were generally undergone episodic positive selection (17 sites), and only one site was under pervasive positive selection. All ZIKV imported into China were Asian lineage and were assigned into two clusters: Venezuela-origin (cluster A) and Samoa-origin cluster (cluster B) with common ancestor from French Polynesia. The time of most recent common ancestors of Cluster A dated to approximately 2013/11 (95% highest posterior density [HPD] 2013/06, 2014/03) and cluster B dated to 2014/08 (95% HPD 2014/02, 2015/01). Cluster B is more variable than Cluster A in comparison with other clusters, but no varied site of biological significance was revealed. ZIKV strains in Southeast Asia countries are independent from strains in America epidemics.

Conclusions: The genetic evolution of ZIKV is conservative. There are two independent introductions of ZIKV into China and China is in danger of autochthonous transmission of ZIKV because of high-risk surrounding areas. Southeast Asia areas have high risk of originating the next large-scale epidemic ZIKV strains.

Keywords: Zika virus; China; Phylogeny; Genetic evolution; Risk assessment

Introduction

Zika virus (ZIKV) is a single-stranded, positive-sense RNA virus in the family _Flaviviridae_, genus _Flavivirus_. While it is transmitted among humans mainly through mosquito-bites (_Aedes aegypti_ and _Aedes albopictus_), it can also be spread through maternal-fetal transmissions, blood transfusions, and sexual activities. In humans, although ZIKV infections typically cause mild and self-limiting febrile illness, it is established that it can also cause severe neurological complications (eg, Guillain-Barré syndrome) and adverse fetal outcomes (eg, microcephaly).[1] ZIKV was first isolated in 1947 from a sentinel rhesus monkey in the Zika forest in Uganda,[2] but it has become a significant global concern in 2015, when it broke out in Brazil,[3] and subsequently spread throughout America. According to the latest update of WHO, as of February 2018, ZIKV has been reported in 86 countries and territories all over the world.[4]

Previous genetic studies have revealed that ZIKV has evolved into three distinct genotypes: West African (Nigerian cluster and Senegal cluster), East African (Uganda cluster), and Asian.[5] However, Asian lineage

Access this article online	
Quick Response Code:	**Website:** www.cmj.org
	DOI: 10.1097/CM9.0000000000000317

is the only genotype known to cause ZIKV outbreaks, whereas no outbreak of African lineage isolates has been reported.[6] While research efforts have mainly focused on ZIKV in Americas, little is known about the genetic relationships among ZIKV strains in Asia, although China and many Southeast Asian countries are in ideal climate conditions of tropical and sub-tropical weather and abundant in competent mosquito vectors for ZIKV transmission. China has experienced a series of imported cases since 2016. From February 2016 to May 2017, 29 ZIKV infection cases were imported into China.[7,8] Singapore has experienced an outbreak of ZIKV in August, 2016,[9] similarly Malaysia, the Philippines, India, Thailand, and Indonesia all reported evidence of ZIKV circulation and transmission before 2016.[4]

Up to now, much remains unknown about ZIKV epidemiology and evolution, in part owing to a lack of genomic data. However, due to the recent epidemics and enabling technologies such as rapid and full-length sequencing, substantial amount of ZIKV genomes can be obtained in a short period of time. China still under high risk of importation and potential localization of ZIKV. So, here we give a detailed description of all the ZIKV isolates and illustrate their introduction path in China. A comprehensive phylogenetic and genetic analysis on isolates in China and surrounding areas was performed. By exploring phylogenetic relationships, selective pressure, and recombination events of ZIKV genomes, much valuable information can be provided for exploring the genetic evolution of ZIKV in China and assessing the risk of ZIKV transmission in the future.

Methods

Sample collection

Selected ZIKV complete/coding sequences were retrieved from the National Center for Biotechnology Information (NCBI) Virus Variation Resource database[10] (hereinafter called NCBI database) as of August 2018. In selecting the ZIKV sequences, all sequences reported from China and only these selected from Zika-endemic areas that are popular Chinese tourist or business destinations were considered. The country list with the number of passengers imported was obtained from China National Bureau of Statistics (http://www.stats.gov.cn/tjsj/), and the top 15 countries with over 5 million passengers arriving China in 2016 were checked for their level of ZIKV-exposing conditions according to WHO Classification Table of ZIKV.[4] Countries that belonged to Categories 1 and 2 were defined as ZIKV-popular areas, and the ZIKV strains from these countries were retrieved from the NCBI database. In addition, ZIKV strains in Brazil, USA, French Polynesia, and Yap Island where ZIKV outbreak has been reported, and those in Venezuela, Samoa, Suriname, and Guatemala where ZIKV in China are imported from, were also collected. ZIKV strains in Uganda, Senegal, and Nigeria were collected as African lineage representatives. Only full-length sequences were included for more informative molecular epidemiological analyses.[11] For convenience of calculations, the number of ZIKV strains was stratify sampled according to the total number of

strains in database. If the total number was less than 5, all strains were included; if the total number was between 5 and 10, 5 strains were included; if the total number of strains was between 10 and 20, half of strains were included; if the total number of strains were more than 20, 10 strains were included. The set of virus sequences included were isolated at different time points. The detail of sampling process can be found in Supplementary Data 1, http://links.lww.com/CM9/A54.

Sequence alignment

All sequences were aligned together using MUltiple Sequence Comparison by Log-Expectation (MUSCLE) implemented in Molecular Evolutionary Genetics Analysis (MEGA) 7.0 (Center for Evolutionary Medicine and Informatics, Philadelphia, PA, USA)[12] and both ends were trimmed by manual editing. Alignment was also carried out by MUSCLE based on codons for selection pressure analysis and the non-synonymous (dN) and synonymous (dS) substitution rates analysis.

Recombination detection

To prevent potential bias during phylogenetic inference due to recombination, sequences of all 88 ZIKV genomes were first analyzed with Recombination Detection Program version 4 (RDP4 program)[13] (University of Cape Town, South Africa) that incorporates seven methods: RDP, Gene Conversion, Chimaera, Maximum Chi-squared, Bootscan, Sister-Scanning, and three Sequences, to uncover evidence for recombination events. Only events with P values ≤ 0.01 that were confirmed by four or more methods were considered to be recombination events.

ML phylogenetic analysis

The suitability of substitution models for our ZIKV alignment was assessed using jModelTest 2.1.10 (GitHub, USA),[14] which performed a statistical model selection procedure based on the Akaike information criterion. This identified the best fitting substitution model General Time Reversible with gamma-distributed rate variation across sites (GTR + Γ) for ML phylogenetic analysis. A ML phylogeny was estimated from this alignment using PhyML 3.0 (ATGC bioinformatics platform, France)[15] under a general time-reversible nucleotide substitution model, with a gamma-distributed among site rate variation (GTR + Γ), as determined by jModelTest2. Statistical support for nodes of the ML phylogeny was assessed using a bootstrap approach with 100 replicates. Support values for all nodes were embedded in the phylogenetic tree files.

Molecular clock and phylogeographic analyses

Temporal evolutionary signal in ML tree was evaluated by using TempEst v1.5.1 (Institute of Evolutionary Biology, USA),[16] which plots sample collection dates against root-to-tip genetic distances obtained from the ML phylogeny tree. A linear regression between sample collection dates and root-to-tip genetic distances obtained from the ML phylogeny indicated that the feasibility of a molecular clock approach. The plot indicated that the data set
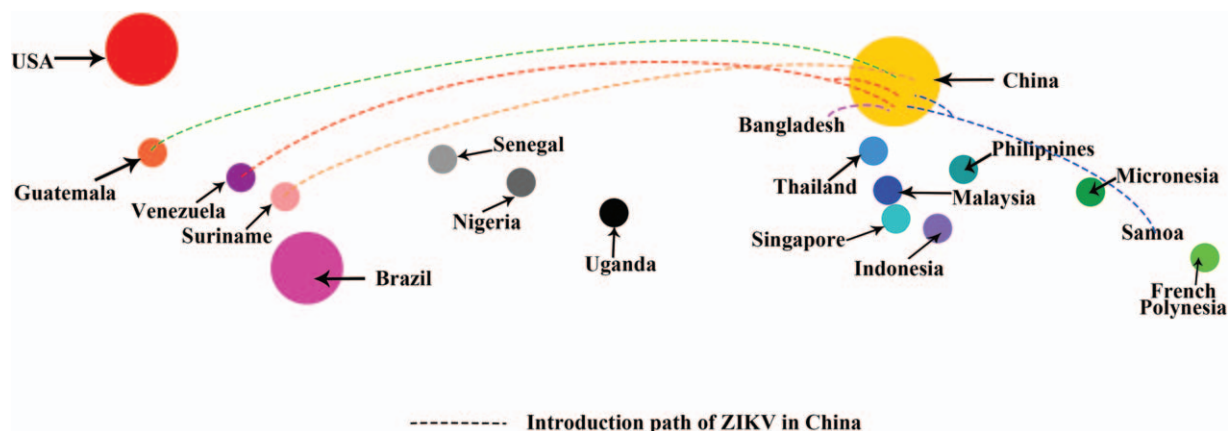
Figure 1: Map where samples were collected in each of the colored areas. The introduction path of ZIKV in China are shown in dotted line. ZIKV: Zika virus.

contained sufficient temporal signal for molecular clock analysis. Molecular clock phylogenies were estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach implemented in BEAST v1.10.1 (Institute of Evolutionary Biology).[17] We computed five independent runs of 100 million MCMC steps, sampling parameters, and trees every 10,000 steps. An uncorrelated lognormal relaxed molecular clock model[18] and a Bayesian skyline coalescent model[19] were used; previous studies have demonstrated this combination to be the best fitting model combination for ZIKV in the Americas.[20,21] In a previous study, different substitution models produced similar results and divergence date estimates are robust among different combinations of prior distributions, molecular clock models, and coalescent models.[11] In each run, a SRD06 substitution model[22] was used, which employs an Hasegawa, Kishino, and Yano (HKY) nucleotide substitution model, a gamma-distribution among site rate variation (HKY + Γ) and a codon position partition (positions [1 + 2] *vs.* position 3). The program Tracer v1.7.1 (Institute of Evolutionary Biology)[23] was used to evaluate MCMC chain convergence and to compute marginal posterior distributions of parameters, after removal of 10% of the chain as burn-in. The program Logcombiner (Institute of Evolutionary Biology) was used to combine and sub-sample posterior tree distributions after a 10% burn-in. TreeAnnotator (Institute of Evolutionary Biology) was used to generate a summary maximum clade credibility (MCC) tree from the posterior distribution of trees (after removal of MCMC burn-in of 10%). The MCC phylogeny tree was drawn using Figtree (Institute of Evolutionary Biology).

Selection pressure analyses

The selection pressures acting on the transmission of ZIKV was inferred by using Datamonkey web server (http://www.datamonkey.org/, USA).[24] To test if positive selection has occurred on a proportion of branches, adaptive branch-site random effects likelihood (aBSREL) method[25] was used. In the scenario where all branches were tested, *P* values at each branch were corrected for multiple testing by using the Holm-Bonferroni correction. Significance was assessed using the likelihood ratio test at a

threshold of $P \leq 0.05$. In addition, we estimated the dN and dS substitution rates on a per-codon site basis for each genome. Sites evolving under positive selection under a proportion of branches were detected by using the mixed effects model of evolution (MEME).[26] Sites under positive selection in the entire phylogeny were detected by using single likelihood ancestor counting (SLAC)[27] and Fast, Unconstrained Bayesian AppRoximation (FUBAR)[28] methods. The significant level was set at $P \leq 0.05$ and posterior probabilities >0.9 in FUBAR.

Amino acid substitutions analyses

MEGA7.0 was used to analyze the variation of amino acid between cluster A and B in China, as well as reference clusters in Singapore, French Polynesia, and Malaysia. The structure of protein was modeled by Swiss-Model.[29] SuperPose Version 1.0 (University of Alberta, Canada) was used to compare the protein structures and get root mean squared deviations (RMSD) value. The structures are considered very similar when RMSD < 1.1 Å.[30]

Results

ZIKV isolates selection

A total of 88 ZIKV sequences were included for analyses [see Supplementary Data S1, S2, http://links.lww.com/CM9/A54, http://links.lww.com/CM9/A55], including 18 from China and 70 from selected countries/regions. It should be mentioned that there are 27 sequences from China in the NCBI database, but nine of them were excluded after checked their sources and sequences from one identical case were removed. The remained 18 sequences are imported from Samoa (9), Venezuela (5), Cambodia (2), Guatemala (1), and Bangladesh (1). The distribution map of the included ZIKV sequences and the introduction path of ZIKV into China are shown in Figure 1.

Recombination events in ZIKV strains

The RDP analysis suggests four recombination events in ZIKV genomes, showed in Table 1. The recombinant

Table 1: Recombination events detected in all ZIKV genome sequences.													
Recombinant	Begin	End	Major parent	Minor parent	R	G	B	M	C	S	T
KF383118|Senegal| 2001/01	1330	1910	LC002520|Uganda| 1947/01	KF383117|Senegal| 1997/01	1.59E-28	1.22E-23	1.29E-27	6.49E-08	9.47E-08	3.04E-07	4.62E-12
KF383116|Senegal| 1968/01	5426	5638	KX198134|Senegal| 1984/01	KF383118|Senegal| 2001/01	2.68E-19	7.41E-17	6.42E-14	2.61E-06	2.46E-06	4.77E-05	4.62E-12
KF383117|Senegal| 1997/01	5090	5576	KX198134|Senegal| 1984/01	LC002520|Uganda| 1947/01	2.56E-19	9.63E-18	6.49E-17	2.71E-08	9.81E-08	2.56E-09	2.31E-12
KF383118|Senegal| 2001/01	8992	9152	LC002520|Uganda| 1947/01	Unknown	3.48E-14	9.07E-10	1.47E-10	1.63E-07	1.17E-06	NS[*]	4.62E-12

[*] No significant P value was recorded for this recombination event using this method. R: RDP; G: GENECONV; B: Bootscan; M: Maxchi; C: Chimaera; S: SiSscan; T: 3Seq; NS: No significant.

strains are all from Senegal and the major parent strains are from Uganda and Senegal. The results suggest that ZIKV African lineage may have experienced several recombination events in the early stage of transmission. However, after it has spread out from Africa and come to form the Asian lineage, the recombination events rarely happen.

Polygenetic analyses

The ML analysis of complete coding regions of the 88 sequences indicates distinct Asian and African lineages. African lineage consists of three independent clusters: Senegal, Uganda, and Nigeria. Asian lineage has caused all the recorded outbreak or endemic circulation of ZIKV in recent decades. All strains imported into China belong to the Asian lineage [Figure S1, http://links.lww.com/CM9/A53]. TempEst plot shows a positive correlation between genetic divergence and sampling time in all sequences ($R^2 = 0.649$) [Figure S2A, http://links.lww.com/CM9/A53]. The correlation becomes stronger within the Asian lineage ($R^2 = 0.996$) [Figure S2B, http://links.lww.com/CM9/A53]. The plot indicated that the data set contained sufficient temporal signal for molecular clock analysis.

The evolutionary rate of ZIKV genomes is estimated to be 8.61×10^{-4} substitutions/site per year (95% highest posterior density [HPD] interval 6.34×10^{-4}– 1.11×10^{-3} substitutions/site per year) by using Bayesian molecular clock phylogenies. As shown in Figure 2, most of the ZIKV Asian lineage isolates that cause recent outbreaks since 2015 are originated from the French Polynesia outbreak lineage in 2013, except Singapore and Thailand. ZIKV isolates imported into China were assigned into two clusters, cluster A and cluster B (posteriori probability [PP] = 1.0). Cluster A is denoted as Venezuela-origin cluster since it consisted of strains all imported from Venezuela into Guangdong, China. Cluster B is denoted as Samoa-origin cluster as most of strains were imported from Samoa into Zhejiang or Guangdong, China, except for two, KY967711 and MF03611 that were imported from Cambodia to Beijing, China. There are three discrete ZIKV isolates that cannot be aggregated into clade: KU744693 from Venezuela to Jiangxi, China; MF593625 from Guatemala to Henan, China, and KY328290 from Bangladesh to Yunnan, China. ZIKV genomes from an outbreak in Singapore 2016 cluster within a distinct, well-supported clade, and are independent from French Polynesia and American strains. The timing of ZIKV introduction into Singapore seems to be parallel with the introduction of ZIKV into French Polynesia. Interestingly, the ZIKV isolates in Brazil and USA are interspersed among isolates from elsewhere in the Americas and do not form distinct clusters, likely implying multiple introductions of ZIKV into the Americas. ZIKV isolates in Venezuela, Suriname, and Guatemala was likely introduce from Brazil. On the other hand, ZIKV isolates in Southeast Asia countries, such as Malaysia, Thailand, Indonesia, and Philippines are not linked to the recent outbreaks in Brazil and USA after 2015. Malaysia clade was perhaps the oldest ZIKV Asian lineages and may be the ancestor of most Asian strains that have recently caused outbreaks.
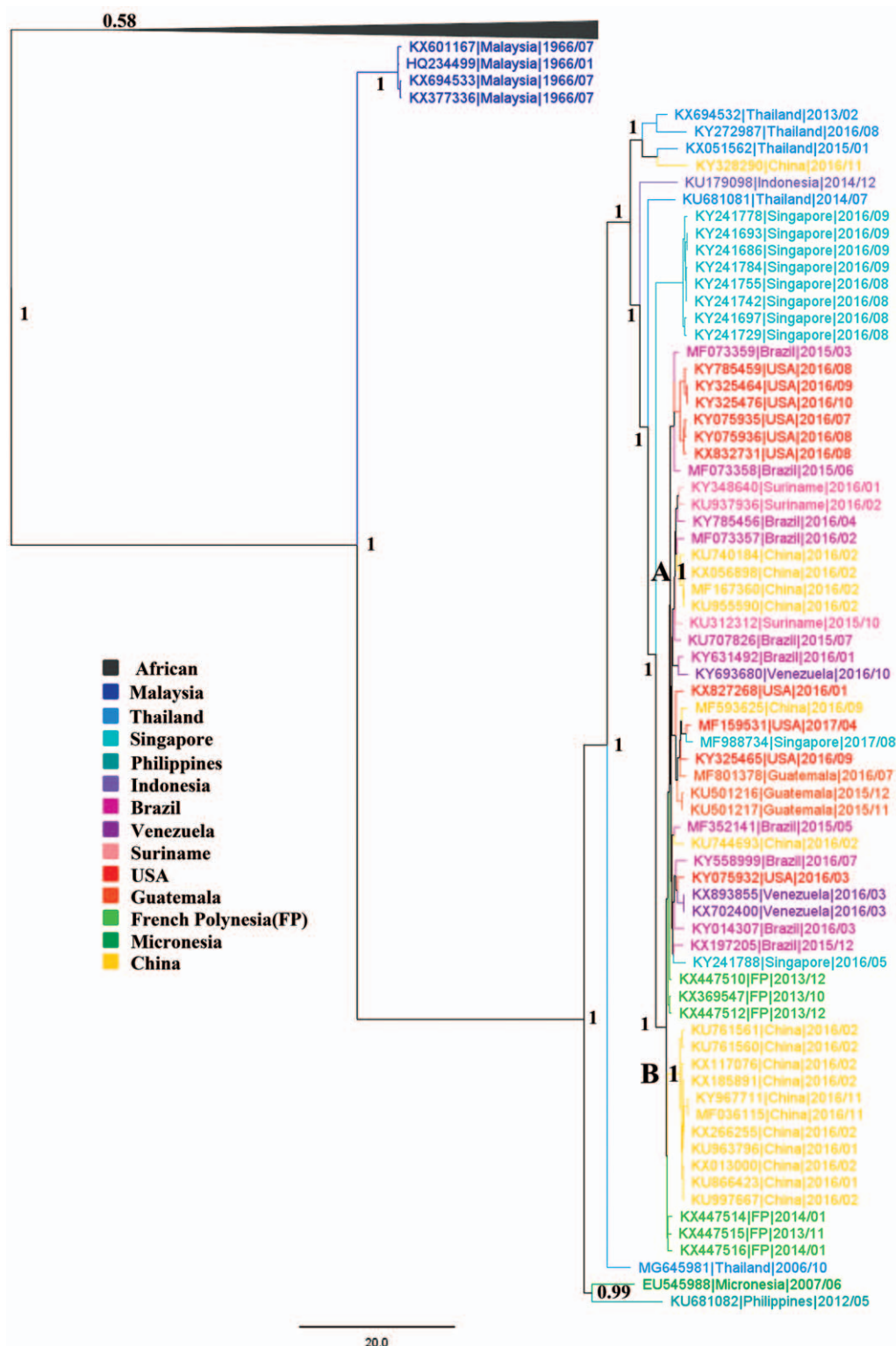
Figure 2: The maximum clade credibility tree of all included Zika virus genome sequence. Node labels are posterior probabilities indicating support for the node.

The time of most recent common ancestors (TMRCA) of Venezuela-origin cluster (cluster A) in China dated to approximately 2013/11 (95% HPD 2013/06, 2014/03) and Samoa-origin cluster (cluster B) dated to 2014/08 (95% HPD 2014/02, 2015/01). This indicates that there were two independent introductions of ZIKV into China.

More specifically, the Venezuela-origin cluster in China was likely transmitted from Brazil in 2013/03, and this is supported by the evidence in Figure 2 suggesting that the isolates in Venezuela were all from Brazil branch. The Samoa-origin cluster in China was introduced from French Polynesia in 2012/05, which appears to be parallel with the

Table 2: Amino acid sites under episodic positive/diversifying selection detected by MEME.

Number	Site	β	P value	Polyprotein location
1	8	35.62	<0.001	Capsid
2	18	163.14	<0.001	Capsid
3	196	2477.23	0.010	PrM
4	693	19.10	0.010	Envelope
5	1143	216.26	0.020	NS1
6	1415	96.53	0.040	NS2B
7	1731	197.22	<0.001	NS3
8	1745	96.28	0.010	NS3
9	1857	145.59	0.020	NS3
10	1872	56.15	<0.001	NS3
11	2086	22.72	0.040	NS3
12	2294	47.34	<0.001	NS4B
13	2356	275.12	<0.001	NS4B
14	2807	64.04	<0.001	NS5
15	2809	1793.20	<0.001	NS5
16	3223	11.41	0.030	NS5
17	3366	72.69	<0.001	NS5

β: Non-synonymous substitution rate at a site for the positive/neutral evolution component. MEME: Mixed effects model of evolution.

introduction of ZIKV in Brazil from French Polynesia. The TMRCA of Singapore clade was estimated in 2015/01 (95% HPD 2014/07, 2015/05), which is before the first ZIKV infection case reported in Singapore (2016/05). The divergence time of Singapore cluster (2011/01, 95% HPD 2009/05, 2012/07) is earlier than French Polynesia cluster (2012/12, 95% HPD 2012/06, 2013/05). ZIKV genomes sampled in Malaysia form a robust monophyletic cluster (PP = 1.0) and share a common ancestor with other ZIKV Asian strains that circulated in recent decades. The Malaysia cluster diverged at a much earlier stage (1957/10, 95% HPD 1949/01, 1964/02) of the Asian lineage.

Selection pressure analyses

Evidence of episodic diversifying selection is detected in three out of all 165 branches in the phylogeny of 88 sequences of ZIKV genome (KU744693_CHINA_2016_02, KF383116_SENEGAL_1968_01, and KF383118_SENEGAL_2001_01) by aBSREL; however, no episodic diversifying selection is identified on internal nodes. As shown in Table 2, there is evidence of episodic positive/diversifying selection at 17 sites as detected by MEME. However, no sites are detected by SLAC. One positively selected site (amino acid site 2449, posterior probability = 0.932, belong to polyprotein NS4B) is detected by FUBAR. The positive selection sites are distributed throughout polyprotein of ZIKV, but most of them located on non-structural proteins, especially on NS3 (5 sites), NS5 (4 sites), and NS4B (3 sites).

Amino acid substitutions in ZIKV clusters

Genetic variation can provide important insights into ZIKV phylogeny and pathogeny. Amino acid substitutions in ZIKV cluster A and cluster B in China, Singapore

Table 3: The sites of amino acid substitution in different ZIKV clusters.

ZIKV clusters	PrM	E	NS1	NS2A	NS5			
Amino acid sites	109	709	1118	1285	2634	2787	3144	3190
China cluster B	N	R	R/Q	I	M	V	S	R
China cluster A	S	K	R	V	V	A	N	K
Singapore	S	K	R	I	M	V	N	K
French Polynesia	S	K	R	I	M	V	N	K
Malaysia	S	K	R	I	T	A	N	K

Cluster A: Venezuela-origin cluster; cluster B: Samoa-origin cluster. Here shown the sites varied between different ZIKV clusters but consistent within each cluster. ZIKV: Zika virus.

cluster, French Polynesia cluster and Malaysia cluster were explored in the analyses. As seen in Table 3, there are eight amino acid sites varied between cluster A and cluster B in China, and they are located throughout ZIKV genome. It is also noted that half (4/8) of the different sites are located in NS5 protein. After compared the two structure models of NS5 protein predicted by Swiss-Model according to two sequences different between these four sites, the structures are similar (RMSD of all atoms = 1.13) [Figure 3]. In comparison with Singapore cluster, French Polynesia cluster, and Malaysia cluster, Samoa-origin cluster (B) in China is more variable than Venezuela-origin cluster (A).

Discussion

Although ZIKV was first isolated in Uganda in Africa, cases of ZIKV infection were only detected sporadically during the next decades. Only 14 human cases were reported before April 2007.[31] The initial ancestor of ZIKV and the potential transmission routes remain elusive mostly due to potential sampling bias[31] and inconclusive adaptation signal of ZIKV from mosquito to human likely as a result of limited availability of isolates.[32] ZIKV has been imported into China since 2016; however, no comprehensive phylogenetic and genetic analyses have been performed based on all ZIKV sequences in China. Since the accuracy of molecular evolution of the virus significantly depends on genome sequences included in the study, we performed a scientific sampling method and suitable sampling size on ZIKV sequences to help fill the gap in understanding ZIKV transmissions in China.

The evolutionary rate of all ZIKV genomes (8.61×10^{-4} substitutions/site per year) is similar to the rate only in Asian lineage (1.05×10^{-3} substitutions/site per year), which is consistent with other study,[33] and the rate is comparable to that of other Flaviviruses.[34,35] Based on the result of polygenetic analysis, ZIKV isolates imported into China are assigned into two clusters, that agrees well with previous study.[33] However, there are three sequences (KU7744693, MF593625, and KY328290) that do not appear to belong to the two clusters. KU7744693 is isolated from the first imported case in China from Venezuela.[36] It was imported into Jiangxi, that is a likely reason why it does not cluster with other strains in
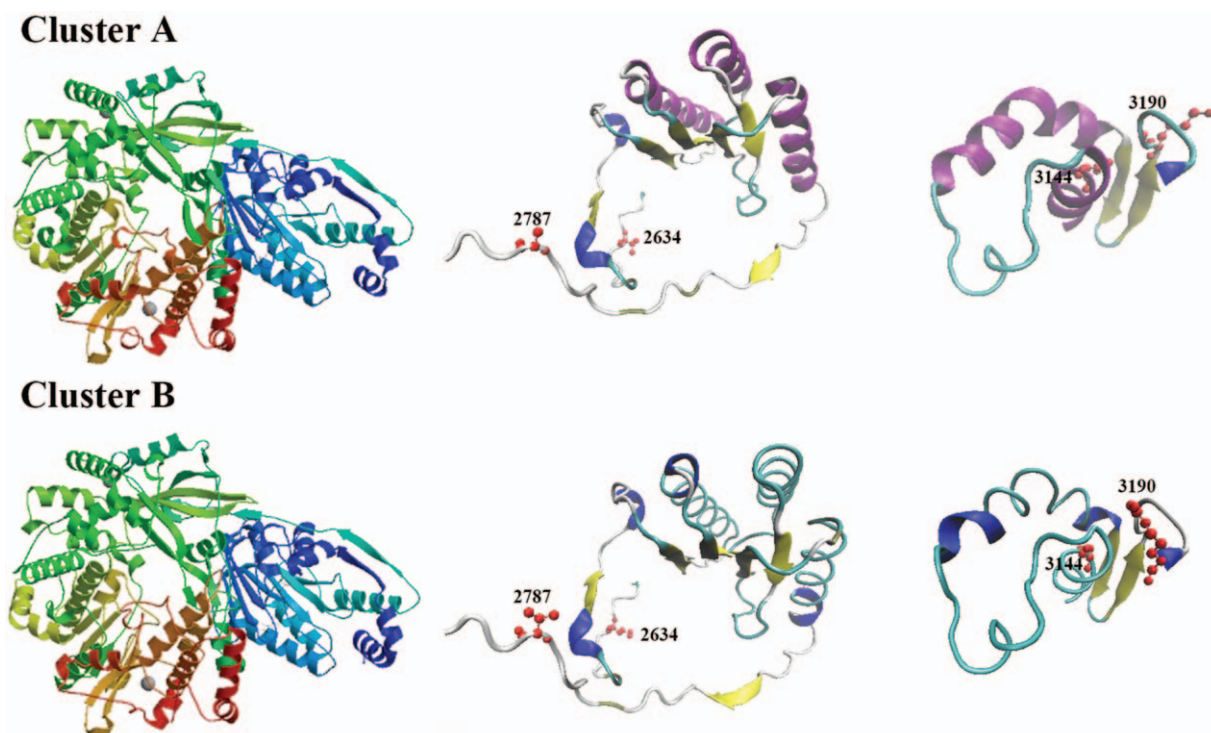
Figure 3: The modeled structure of NS5 protein with amino acid substitutions between two clusters in China.

Venezuela-origin cluster which were all imported into Guangdong. MF593625 is a strain imported from Guatemala and it shares a common ancestor with strains in Guatemala. It is remarkable to note that that strain KY328290 was originally imported into Yunnan, China from Bangladesh in 2014.[37] It is undetected until retesting the ice-stored serum samples retrospectively in 2016 and reported as the first recorded ZIKV imported case in China. KY328290 share a common ancestor with the ZIKV strain in Thailand in September 2010. Although the case is imported from Bangladesh, no ZIKV strain in Bangladesh was included in the NCBI database. It was confirmed as an Asian-lineage transition-state intermediate between low-prevalence and explosive pandemic strains. This strain indicates that China may have imported ZIKV long before 2016, so, the risk of autochthonous transmission of ZIKV in China has increased.

ZIKV was first reported in Asia in Malaysia in 1966[38] and it may be the ancestor of all ZIKV Asian strains. While ZIKV is widespread throughout Thailand, very few strains were reported. This may reflect the limited capacity of ZIKV detection in Thailand.[39] The outbreak lineage in Singapore probably originated within Asia.[9] It is consistent with our polygenetic analyses that outbreak in Singapore in 2016 is not linked to the South American epidemics. There are two divergent ZIKV strains in Singapore. The MF988734 isolate is a strain imported into Singapore from Cuba, and it is distinctly different from autochthonous strains circulating in Singapore.[40] The KY241788 isolate is from the first confirmed case of ZIKV infection in Singapore in May 2016, who had traveled from Brazil. ZIKV in French Polynesia was

introduced from Southeast Asia and then transmitted into South America causing pandemic spread of Zika disease. The strain that responsible for the epidemic of ZIKV in Yap Island in 2007 and the Cambodian cases were also most likely originated in Southeast Asia.[5] Therefore, it is likely that ZIKV may have established latent circulation in Southeast Asia without causing any noticeable outbreak, while still remains as a potential source of strains that can cause outbreaks, which should be considered seriously.

Selection pressure analyses in all included ZIKV sequences shows that three positive selections have occurred on all branches, but no positive selection has been identified on internal nodes. However, this may be not interpreted as ZIKV strains experienced positive selection in the three branches. Because tip branches are measured over a short sampling period, therefore may include a higher proportion of mutations that have not yet been removed through purifying selection. MEME aims to detect the individual sites under episodic positive/diversifying selection. But SLAC and FUBAR were used to detect sites under pervasive positive selection. The different result between MEME and SLAC/FUBAR indicates that ZIKV genomes were undergone generally episodic positive selection but rarely pervasive positive selection. The result shows that ZIKV genome is generally conserved, which is in line with other study.[41] Although there are eight amino acid substitutions in China cluster A and B, none of them were reported of significant biological influence. The amino acid substitution S109N, K709R, and N3144S were tested and showed non-significant enhance of ZIKV infection.[42] M2634V substitution was reported of negligible contribu-

tion to ZIKV pathogenesis.[43] The biological significance of other substitutions remains to be clarified.

Based on the polygenetic results in our study, we can conclude that the selection of sequences via scientific sampling in this study is a well representative of all ZIKV sequences. The computational redundancy can be lessened dramatically by scientifically sampling of sequences and without reducing accuracy. It is reasonable to included only one of these sequences isolated from one same case, although these samples were sequenced by different institutes using different methods, few differences were observed in the genomes.[33] But there are spaces for improvements to this study in the future. On one hand, the number of sequences in some countries/regions is too little to draw any conclusions. For example, only one sequence was reported in each of the countries/regions, such as the Philippines, Indonesia, and Micronesia and none was reported in India and Samoa. This may reflect differences in surveillance intensity among different countries. On the other hand, the genomes isolated in China used in this study were all retrieved from NCBI database; however, it may be better that all samples are re-collected and re-sequenced to provide more information and avoid possible confounding factors that may affect the results.

In conclusion, the genetic evolution of ZIKV is conservative. There are two independent introductions of ZIKV into China and China are in danger of autochthonous transmission of ZIKV because of high-risk surrounding areas. Southeast Asia areas have high risk of originating the next large-scale epidemic ZIKV strains.

Conflicts of interest

None.

References

1. Petersen LR, Jamieson DJ, Powers AM, Honein MA. Zika virus. N Engl J Med 2016;374:1552–1563. doi: 10.1056/NEJMra1602113.
2. Dick GW, Kitchen SF, Haddow AJ. Zika virus. I. Isolations and serological specificity. Trans R Soc Trop Med Hyg 1952;46:509–520. doi: 10.1016/0035-9203(52)90042-4.
3. Campos GS, Bandeira AC, Sardi SI. Zika virus outbreak, Bahia, Brazil. Emerg Infect Dis 2015;21:1885–1886. doi: 10.3201/eid2110.150847.
4. World Health Organization. Zika Virus Classification Tables; 9 March 2018. Available from: http://apps.who.int/iris/bitstream/handle/10665/260419/zika-classification-15Feb18-eng.pdf;jsessionid=B67AE00B5428988C6172293700B1296E?sequence=1. [Accessed 2018-12-04].
5. Haddow AD, Schuh AJ, Yasuda CY, Kasper MR, Heang V, Huy R, et al. Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. PLoS Negl Trop Dis 2012;6:e1477. doi: 10.1371/journal.pntd.0001477.
6. Ramaiah A, Dai L, Contreras D, Sinha S, Sun R, Arumugaswami V, et al. Comparative analysis of protein evolution in the genome of pre-epidemic and epidemic Zika virus. Infect Genet Evol 2017;51:74–85. doi: 10.1016/j.meegid.2017.03.012.
7. Zhou PC, Yang ZJ, Wang J. Enlightenment of handling the first imported Zhaika virus infection case in Jiangsu province (in Chinese). China Inspection and Quarantine Times 2017-05-25.
8. Sun J, Wu D, Zhong H, Guan D, Zhang H, Tan Q, et al. Returning ex-patriot Chinese to Guangdong, China, increase the risk for local transmission of Zika virus. J Infect 2017;75:356–367. doi: 10.1016/j.jinf.2017.07.001.
9. Ho ZJM, Hapuarachchi HC, Barkham T, Chow A, Ng LC, Lee JMV, et al. Outbreak of Zika virus infection in Singapore: an epidemiological, entomological, virological, and clinical analysis. Lancet Infect Dis 2017;17:813–821. doi: 10.1016/S1473-3099(17)30249-9.
10. Hatcher EL, Zhdanov SA, Bao Y, Blinkova O, Nawrocki EP, Ostapchuck Y, et al. Virus variation resource - improved response to emergent viral outbreaks. Nucleic Acids Res 2017;45:D482–D490. doi: 10.1093/nar/gkw1065.
11. Faria NR, Azevedo R, Kraemer M, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: early epidemiological and genetic findings. Science 2016;352:345–349. doi: 10.1126/science.aaf5036.
12. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 2016;33:1870–1874. doi: 10.1093/molbev/msw054.
13. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol 2015;1:v3. doi: 10.1093/ve/vev003.
14. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 2012;9:772. doi: 10.1038/nmeth.2109.
15. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 2010;59:307–321. doi: 10.1093/sysbio/syq010.
16. Rambaut A, Lam TT, Max CL, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol 2016;2:w7. doi: 10.1093/ve/vew007.
17. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 2012;29:1969–1973. doi: 10.1093/molbev/mss075.
18. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol 2006;4:e88. doi: 10.1371/journal.pbio.0040088.
19. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 2005;22:1185–1192. doi: 10.1093/molbev/msi103.
20. Metsky HC, Matranga CB, Wohl S, Schaffner SF, Freije CA, Winnicki SM, et al. Zika virus evolution and spread in the Americas. Nature 2017;546:411–415. doi: 10.1038/nature22402.
21. Grubaugh ND, Ladner JT, Kraemer M, Dudas G, Tan AL, Gangavarapu K, et al. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. Nature 2017;546:401–405. doi: 10.1038/nature22400.
22. Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Mol Biol Evol 2006;23:7–9. doi: 10.1093/molbev/msj021.
23. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using tracer 1.7. Syst Biol 2018;67:901–904. doi: 10.1093/sysbio/syy032.
24. Pond SL, Frost SD. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 2005; 21:2531–2533. doi: 10.1093/bioinformatics/bti320.
25. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky PS. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol 2015;32:1342–1353. doi: 10.1093/molbev/msv022.
26. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky PS. Detecting individual sites subject to episodic diversifying selection. PLoS Genet 2012;8:e1002764. doi: 10.1371/journal.pgen.1002764.
27. Kosakovsky PS, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 2005;22:1208–1222. doi: 10.1093/molbev/msi105.

28. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky PS, et al. FUBAR: a fast, unconstrained Bayesian approximation for inferring selection. Mol Biol Evol 2013;30:1196–1205. doi: 10.1093/molbev/mst030.

29. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 2018;46:W296–W303. doi: 10.1093/nar/gky427.

30. Maiti R, Van Domselaar GH, Zhang H, Wishart DS. SuperPose: a simple server for sophisticated structural superposition. Nucleic Acids Res 2004;32:W590–W594. doi: 10.1093/nar/gkh477.

31. Faye O, Freire CC, Iamarino A, Faye O, de Oliveira JV, Diallo M, et al. Molecular evolution of Zika virus during its emergence in the 20(th) century. PLoS Negl Trop Dis 2014;8:e2636. doi: 10.1371/journal.pntd.0002636.

32. Wang L, Valderramos SG, Wu A, Ouyang S, Li C, Brasil P, et al. From mosquitos to humans: genetic evolution of Zika virus. Cell Host Microbe 2016;19:561–565. doi: 10.1016/j.chom.2016.04.006.

33. Sun J, Wu , Zhong H, Guan D, Zhang H, Tan Q, et al. Returning expatriot Chinese to Guangdong, China, increase the risk for local transmission of Zika virus. J Infect 2017;75:356–367. doi: 10.1016/j.jinf.2017.07.001.

34. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, et al. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. Proc Natl Acad Sci U S A 2012;109:15066–15071. doi: 10.1073/pnas.1206598109.

35. Sall AA, Faye O, Diallo M, Firth C, Kitchen A, Holmes EC. Yellow fever virus exhibits slower evolutionary dynamics than dengue virus. J Virol 2010;84:765–772. doi: 10.1128/JVI.01738-09.

36. Liu L, Wu W, Zhao X, Xiong Y, Zhang S, Liu X, et al. Complete genome sequence of Zika virus from the first imported case in Mainland China. Genome Announc 2016;4:e00291–e00316. doi: 10.1128/genomeA.00291-16.

37. Wang B, Liang Y, Lu Y, Zhang L, Li Y, Song Y, et al. The importation of the phylogenetic-transition state of Zika virus to China in 2014. J Infect 2018;76:106–109. doi: 10.1016/j.jinf.2017.09.002.

38. Marchette NJ, Garcia R, Rudnick A. Isolation of Zika virus from Aedes aegypti mosquitoes in Malaysia. Am J Trop Med Hyg 1969;18:411–415. doi: 10.4269/ajtmh.1969.18.411.

39. Buathong R, Hermann L, Thaisomboonsuk B, Rutvisuttinunt W, Klungthong C, Chinnawirotpisan P, et al. Detection of Zika virus infection in Thailand 2012–2014. Am J Trop Med Hyg 2015;93:380–383. doi: 10.4269/ajtmh.15-0022.

40. Tan CH, Tan LK, Hapuarachchi HC, Lai Y, Wong P, Yap G, et al. Viral and antibody kinetics, and mosquito infectivity of an imported case of Zika fever due to asian genotype (American Strain) in Singapore. Viruses 2018;10:44. doi: 10.3390/v10010044.

41. Marquez-Jurado S, Nogales A, Avila-Perez G, Iborra F, Martínez-Sobrido L, Almazán F. An alanine-to-valine substitution in the residue 175 of Zika virus NS2A protein affects viral RNA synthesis and attenuates the virus in vivo. Viruses 2018;10:547. doi: 10.3390/v10100547.

42. Yuan L, Huang XY, Liu ZY, Zhang F, Zhu XL, Yu JY, et al. A single mutation in the prM protein of Zika virus contributes to fetal microcephaly. Science 2017;358:933–936. doi: 10.1126/science.aam7120.

43. Zhao F, Xu Y, Lavillette D, Zhong J, Zou G, Long G. Negligible contribution of M2634V substitution to ZIKV pathogenesis in AG6 mice revealed by a bacterial promoter activity reduced infectious clone. Sci Rep 2018;8:10491. doi: 10.1038/s41598-018-28890-0.