

Genomic variants and molecular epidemiological characteristics of dengue virus in China revealed by genome-wide analysis

Bangyao Sun^{1,†}, Meng Xu^{2,3,4,5,†}, Lijia Jia^{2,3,†}, Haizhou Liu^{2,3}, Aixin Li¹, Lixia Hui¹, Zhitao Wang^{6,*}, Di Liu^{2,3,*}, Yi Yan^{7,8,*}

¹School of Medical Laboratory, Shandong Second Medical University, Baotong West Street 7166#, Weifang 261053, China

²CAS Key Laboratory of Special Pathogens and Biosafety, Wuhan Institute of Virology, Chinese Academy of Sciences, Jinlong Street 262#, Wuhan 430207, China

³Computational Virology Group, Center for Bacteria and Viruses Resources and Bioinformatics, Wuhan Institute of Virology, Chinese Academy of Sciences, Jinlong Street 262#, Wuhan 430207, China

⁴Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Kaiyuan Avenue 190#, Guangzhou 510530, China

⁵University of Chinese Academy of Sciences, Yuquan Road 19#, Beijing 100049, China

⁶School of Life Science and Technology, Shandong Second Medical University, Baotong West Street 7166#, Weifang 261053, China

⁷Department of Respiratory Medicine, Wuhan Children's Hospital, Tongji Medical College, Huazhong University of Science and Technology, Hong Kong Road 100#, Wuhan 430015, China

⁸Pediatric Respiratory Disease Laboratory, Institute of Maternal and Child Health, Wuhan Children's Hospital, Tongji Medical College, Huazhong University of Science and Technology, Hong Kong Road 100#, Wuhan 430015, China

[†]These authors contributed equally to this article.

*Corresponding authors. Yi Yan, E-mail: yanyi_1996@163.com; Di Liu, E-mail: liud@wh.iov.cn; Zhitao Wang, E-mail: wangzhitao@sdsu.edu.cn

Abstract

Since its first academic record in 1978, dengue epidemics have occurred in all provinces of China, except Xizang. The epidemiological and molecular features of the whole genome of dengue virus (DENV) have not yet been completely elucidated, interfering with prevention and control strategies for dengue fever in China. Here, we obtained 553 complete genomes of the four serotypes of DENV (DENV1–4) isolated in China from the GenBank database to analyze the phylogeny, recombination, genomic variants, and selection pressure and to estimate the substitution rates of DENV genomes. Phylogenetic analyses indicated that DENV sequences from China did not cluster together and were genetically closer to those from Southeast Asian countries in the maximum likelihood trees, indicating that DENV was not endemic in China. Thirty intra-serotype recombinant sequences were identified for DENV1–4, with the highest frequency in DENV4. Selection pressure analyses revealed that 13 codons under positive selection were located in the C, NS1, NS2A, NS3, and NS5 proteins. For DENV1 to DENV3, the substitution rates evaluated in this study were 9.23×10^{-4} , 7.59×10^{-4} , and 7.06×10^{-4} substitutions per site per year, respectively. These findings improve our understanding of the evolution of DENV in China.

Keywords: dengue virus in China; complete genome; phylogenetic analysis; recombination detection; selection pressure; single nucleotide polymorphism (SNP); substitution rate

Introduction

Dengue fever (DF) is a mosquito-borne (*Aedes aegypti* or *Aedes albopictus*) disease distributed in tropical and subtropical regions (Wilder-Smith et al. 2019). DF is caused by dengue virus (DENV), which belongs to the genus *Flavivirus* of the family *Flaviviridae*. The DENV genome is ~11 kb in length and is divided into four antigenically distinct serotypes (DENV1–4) (Chen and Vasilakis 2011). Most dengue infections are asymptomatic or mild, and ~1%–2% of dengue cases progress to dengue hemorrhagic fever/dengue shock syndrome, with a case fatality rate of 5% (Kyle and Harris 2008, Tayal et al. 2023). DENV is endemic in many countries of Southeast Asia, the Americas, and the Western Pacific, and its prevalence has increased in recent years (Xu et al. 2017). According to World

Health Organization (WHO) reports, ~100–400 million infections occur annually worldwide (WHO, 2021), posing a significant burden on the global economy and health (Shepard et al. 2016, Xu et al. 2022).

In China, a dengue outbreak in Foshan, Guangdong Province, was first academically recorded in 1978 and has since occurred frequently in six provinces (Guangdong, Guangxi, Hainan, Yunnan, Fujian, and Zhejiang) in southern China (Lin et al. 2020). Due to climate change, dengue epidemics have become increasingly prevalent in northern China in recent years (Yue et al. 2019). To date, dengue outbreaks have been reported in all provinces of China except Xizang, with more than 740 000 infected cases. Currently, there are no effective antiviral drugs available for the

treatment of dengue. Dengvaxia, the only vaccine available for dengue prevention, has not yet been approved for use in China (Wu et al. 2022). China's territory spans tropical, subtropical, and temperate zones with over 1.4 billion people; therefore, elucidation of the molecular epidemiological and evolutionary characteristics of the DENV genome is of great significance for understanding the epidemic status and guiding surveillance and prevention measures for DF in China and worldwide (Wu et al. 2022).

In previous studies, the distribution characteristics of dengue infections in China (Yue et al. 2019, Lin et al. 2020, Wu et al. 2022), seroprevalence and risk assessment of DENV infections in selected provinces (Sang et al. 2021, Wang et al. 2021, Cui et al. 2022, Liu et al. 2022, Zhang et al. 2023), and epidemiological and evolutionary characteristics of dengue in high-risk provinces based on the E gene of DENV have been widely reported (Wu et al. 2011, Sang et al. 2015, Du et al. 2021, Yao et al. 2021, Zhang et al. 2021, Meng et al. 2023, Sun et al. 2023). Viral evolutionary characteristics at the population level could be revealed by single-nucleotide polymorphisms (SNPs). Utilizing complete viral genomes is important to accurately reconstruct the phylogeny and molecular epidemiology (Zhu et al. 2022, Zh et al. 2024). Compared with the E gene, complete genomes can provide more comprehensive and accurate information on the genomic variants of DENV in China to enhance interventions for DENV infection. However, the genomic variants of the whole genomes of DENV1–4 in China remain unknown. In this study, we downloaded all the complete genomes of DENV isolated in China from the GenBank database and investigated their phylogenetics, recombination, selection pressure, and genomic variants. Furthermore, substitution rates of DENV1–3 were estimated using the full-length DENV genome. Therefore, this study provides new insights into the evolutionary characteristics of DENV in China.

Materials and methods

Dataset

All DENV1–4 full-length genomes isolated from China were downloaded from the GenBank database (up to 31 March 2022); those without the year of isolation were excluded. Next, each DENV1–4 genome was subjected to online BLASTN analysis, with the top 50 selected hits. After eliminating duplicate sequences or sequences with poor sequencing quality, four datasets (DENV1–4) were generated for subsequent analyses. Among the four datasets, the numbers of DENV genomes from China and other countries were 244 and 874 for DENV1, 215 and 735 for DENV2, 68 and 372 for DENV3, and 26 and 245 for DENV4, respectively, with their GenBank accession numbers listed in Table S1. UTR sequences were removed from all analyses because of their inconsistent lengths.

Phylogenetic analysis

Before phylogenetic analysis, recombination detection of the four datasets was performed using Recombination Detection Program 4 (RDP4) (Martin et al. 2015), and the recombinant sequences isolated from other countries were excluded from the datasets. The maximum likelihood (ML) method was used to construct phylogenetic trees and implemented in RAXML v8.2.12 (Stamatakis 2014). The bootstrap value was set to 1000, and general time reversible (GTR) was used as the nucleotide substitution model. Four published sequences (GenBank accession numbers JQ920481, KX812530, AB189121, and AY037116) were selected as outgroups for DENV1, DENV2, DENV3, and DENV4, respectively, and were removed for subsequent presentation.

Recombination analysis

RDP4 (Martin et al. 2015) was used to perform recombination detection using default parameters. Seven methods, including RDP (Martin et al. 2015), GENECONV (Padidam et al. 1999), BootScan (Martin et al. 2005), MaxChi (Smith 1992), Chimaera (Posada and Crandall 2001), SiScan (Gibbs et al. 2000), and 3Seq (Boni et al. 2007) were selected to perform the recombination analysis of DENV genomic sequences, and P-values < .001 were considered a positive signal. A recombination event was identified when a recombinant signal was detected by at least three methods. Finally, 30 recombinant sequences were detected and their detailed detection information is presented in Table S2.

Substitution analysis of DENV genomes

SNPs from Chinese DENV genomes were gathered using a homemade Practical Extraction and Report Language (Perl) script (<https://github.com/zeroliu/bioutils/tree/master/snp>), and the consensus sequences of DENV1–4 datasets were obtained using the European Molecular Biology Open Software Suite (EMBOSS) package (Rice et al. 2000).

Selection pressure analysis

Recombinant sequences detected by RDP4 were excluded from the four datasets (DENV1–4), and the remaining sequences were used for the selection pressure analysis with stop codons removed. Four methods [mixed effects model of evolution (MEME) (Murrell et al. 2012), single likelihood ancestor counting (SLAC) (Kosakovsky Pond and Frost 2005), fixed effects likelihood (FEL) (Bartolucci et al. 2016), and fast unbiased Bayesian approximation, FUBAR, (Murrell et al. 2013)] were selected using the HyPhy (Hypothesis Testing using Phylogenies) package (<http://hyphy.org/>) with default values for the significance levels (as for SLAC, FEL, MEME, P-value = .1; the posterior probability of FUBAR = .9). When at least three methods showed the dN/dS value (nonsynonymous substitution rate/synonymous substitution rate) > 1, the site was identified as a positive selection site.

Nucleotide substitution rates of DENV genome

After sequence alignment, the mean nucleotide substitution rates of the DENV1, DENV2, and DENV4 genomes were estimated using the Bayesian Markov chain Monte Carlo method implemented in BEAST1.10.4 (Suchard et al. 2018), with a relaxed clock model and the GTR + I + G4 model. The mean nucleotide substitution rate of the DENV3 genome was estimated using BEAST2 (Remco et al. 2014), with a strict clock model and the GTRGAMMA substitution model. A total of 800, 2400, 500, and 800 million steps were run for Bayesian Markov chain Monte Carlo analyses of the DENV1, DENV2, DENV3, and DENV4 genomic alignments, respectively, with 10% burn-in. The trees and other parameters were sampled at every 10 000 steps. The detailed parameters are listed in Table S4.

Results

Distribution of DENV genomes from China

In this study, we included 553 complete DENV genomes isolated in China as of 31 March 2022. DENV1–2 genomes accounted for 83.0% of the total (Table 1; DENV1 = 244, DENV2 = 215, DENV3 = 68, and DENV4 = 26). In terms of temporal distribution, only 18 full-length DENV genomes were available before 2011; from 2012 to 2019, 539 full-length DENV genomes were published, but no sequences were available from 2020 to March 2022 (Table 1).

Table 1. The number of whole genomes of DENV1–4 from China in indicated years.

Year	DENV1	DENV2	DENV3	DENV4
Before 2000		2		
2000				
2001		1		
2002	1			
2003				
2004	1			
2005	1			
2006	2			
2007	1		1	
2008				
2009			2	
2010		2		3
2011	1			
2012	2	1	1	1
2013	10	5	22	2
2014	72	12	1	
2015	57	65	10	11
2016	33	15	6	2
2017	18	84	3	
2018	10	11	4	6
2019	35	17	18	1
2020				
2021				
2022				
Total	244	215	68	26

Of these 553 genomes, 12 had undetermined collection provinces, and the rest were distributed in 10 of the 34 provincial-level administrative districts in China. The genomes of DENV1, DENV2, DENV3, and DENV4 were distributed in seven, six, four, and two provinces of China, respectively (Fig. 1). The number of DENV genomes generally increased from the north to south (Fig. 1). The number of DENV genomes in Henan, Hebei, Jiangsu, Hubei, Fujian, Guangxi, and Hainan provinces was <10, whereas that in Zhejiang, Yunnan, and Guangdong provinces was >40. Only the genomes from Yunnan and Guangdong provinces cover simultaneously DENV1–4. Most DENV genomes were distributed in Guangdong Province, accounting for 70.5% (390/553; Fig. 1).

Phylogenetics of DENV genomes from China

Phylogenetic analysis was performed to reveal the phylogenetic relationships and epidemic characteristics of the DENV1–4 genomes from China. In general, the genomes of DENV1–4 from different years or provinces did not cluster together in the ML trees, and most were genetically close to those from Southeast Asian countries, suggesting a possible transmission relationship between China and Southeast Asian countries during dengue outbreaks (Figs. 2 and 3).

The distribution of DENV genomes along ML trees was characterized differently in different provinces in China. In Guangdong, the province with the worst DF epidemics in China (Yue et al. 2021), the DENV1–4 genomes from different years presented diverse lineages and were not clustered together in the ML trees. Compared to the DENV1 genomes, the most prevalent dengue serotype in Guangdong Province (Ma et al. 2021), the DENV2–4 genomes were more genetically distant from those from other provinces in China on ML trees (Figs. 2 and 3). Yunnan is another province in China where dengue epidemics have been notable (Yue et al. 2021). DENV2 genomes from Yunnan were grouped into separate clusters (Fig. 2b), and these sequences were mainly

derived from the 2015 dengue outbreak in Xishuangbanna, Yunnan Province (Jiang et al. 2018). The DENV3–4 genomes from Yunnan were genetically closer to those from overseas than to those from Guangdong (Fig. 3), suggesting that the dengue epidemics in Yunnan involved independent transmission events. The DENV2 and DENV3 genomes, originating from the dengue epidemics in Hangzhou, Zhejiang Province, in 2017 (Yan et al. 2018) and the center of Henan Province in 2013 (Huang et al. 2014), respectively, were clustered together and were all genetically close to sequences from Southeast Asian countries (Figs. 2b and Fig. 3a). DENV genomes from other provinces in China were scattered across the ML trees.

Recombination of DENV genomes from China

In this study, 30 intra-serotype recombinant sequences were detected (9 in DENV1, 13 in DENV2, 2 in DENV3, and 6 in DENV4), with no inter-serotype recombinant events (Fig. 4a and b). Among them, 15, 9, and 3 recombinant sequences were obtained from Guangdong, Yunnan, and Zhejiang provinces, respectively, and the locations of the remaining three sequences were unknown. Recombinant sequences from the Guangdong Province were included in DENV1–4 (Fig. 4b). The highest number and proportion of recombinant sequences were found for DENV2 (13/215) and DENV4 (6/26), respectively (Fig. S1A). Recombination of DENV4 genomes was found in all genes except 2K and NS4B, whereas recombination of DENV3 genomes was detected only in the NS3 and NS5 genes. Recombination in the DENV1 and DENV2 genomes occurred in seven and six genes, respectively (Fig. 4a and Fig. S1B).

We then examined the recombinant number and frequency of DENV genes. Our findings showed that recombination events were detected in all genes, except 2K and NS4B, with most recombination events detected in the NS5 gene (Fig. S1B). Recombination events of the NS3 and NS5 genes were identified for DENV1–4, and recombination of the NS4A gene occurred only in DENV4. Recombination events in six genes (C, M, E, NS1, NS2A, and NS2B) were detected in at least two dengue serotypes (Fig. S1B). As gene lengths and sequence numbers varied among the DENV1–4 genomes, we normalized the number of recombinant events. Our results indicated that the number of recombination events per gene decreased after normalization treatment and more recombination events occurred in the C, M, NS2A, and NS2B genes than in the other genes (Fig. 4c). The highest proportion of recombinants was observed for DENV4 (Fig. 4c).

Nucleotide substitutions and selection pressure of DENV genomes from China

Viral evolutionary analysis was performed based on SNPs in the viral genome, particularly nonsynonymous substitutions. The molecular characteristics of DENV genomes from China were investigated using a custom Perl script to call SNPs (Sun et al. 2020). Our investigation showed that there were far more synonymous than nonsynonymous substitutions in each dengue serotype (Fig. S2A). We then plotted the distribution of nonsynonymous DENV1–4 substitutions by province (Fig. 5a; synonymous substitutions are not shown). In agreement with the results of phylogenetic analyses, different distribution characteristics of nonsynonymous SNPs were observed among the DENV genomes from Yunnan, Zhejiang, and Guangdong.

We further investigated substitutions within each gene of the DENV genome. Both synonymous and nonsynonymous substitutions were detected in all the genes. In general, the longer the

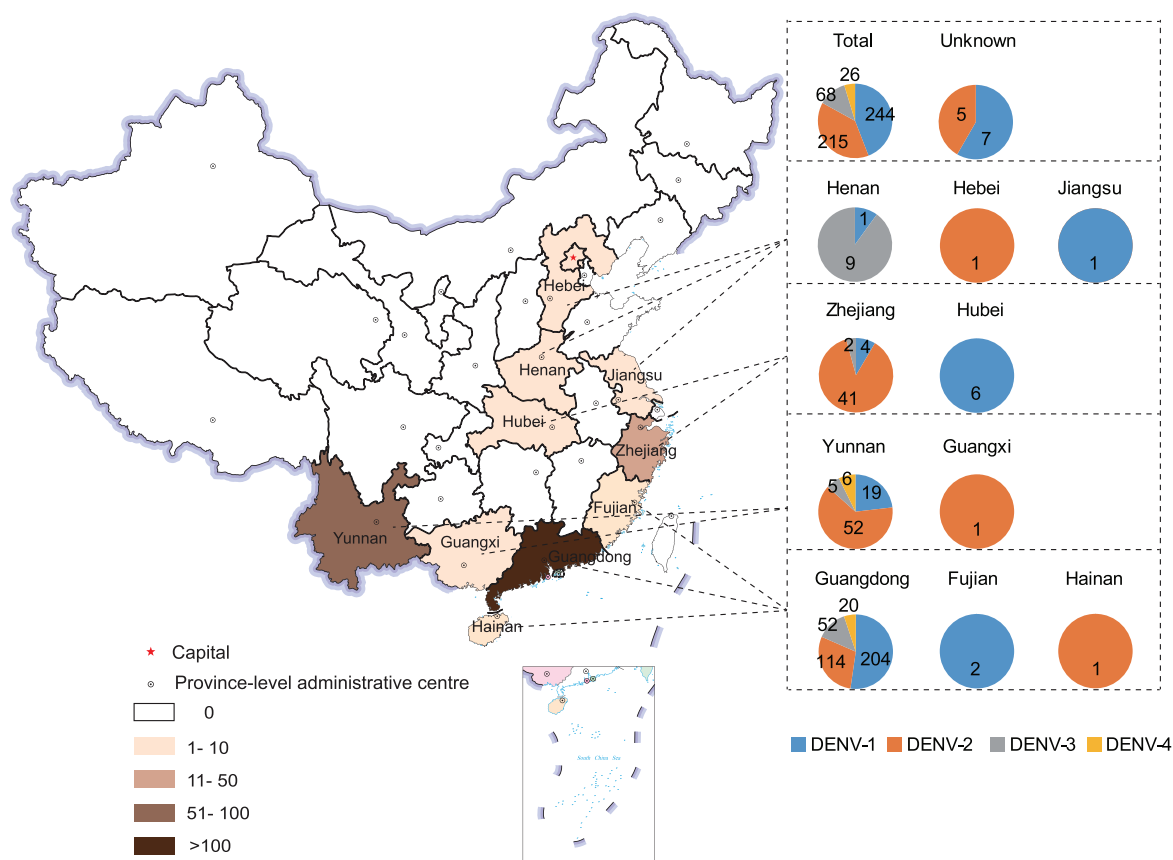


Figure 1. Geographical distribution of DENV genomes in China. The pie charts represent the serotype distribution of the DENV1–4 full-length genomes in indicated provinces. The “Unknown” means that the location of the genome is unknown. The map image is from the standard map copyright-free service system in China (<http://bzdt.ch.mnr.gov.cn/>).

gene sequence, the greater the number of accumulated substitutions (Fig. S2B). Next, the normalized SNPs per gene of the DENV genome were gathered and compared with those of the entire coding sequence (CDS). Our results indicated that no significant difference was found in either synonymous or nonsynonymous substitutions (Fig. 5b, $P > .01$).

The estimated results of the codon-specific selection analysis showed that most codons were under negative selection and only 13 codons were under positive selection (6 in DENV1, 3 in DENV2, 1 in DENV3, and 3 in DENV4, Table 2), suggesting that each gene of DENV1–4 was subjected to similar selective pressures. Of these 13 positive selection codons, eleven were in the nonstructural protein coding region and only two were in the structural protein coding region. The highest number of positive selection codons was detected in the NS5 protein, and no positive codons were found in the M, E, NS2B, NS4A, 2K, or NS4B proteins (Table 2).

Additionally, we estimated the mean nucleotide substitution rates of DENV genomes in China. For DENV1 to DENV3, they were 9.23×10^{-4} substitutions/site/year (s/s/y) (95% highest posterior density, 95% HPD, 6.86×10^{-4} to 1.17×10^{-3}), 7.59×10^{-4} s/s/y (95% HPD, 3.90×10^{-4} to 9.89×10^{-4}), and 7.06×10^{-4} s/s/y (95% HPD, 5.76×10^{-4} to 8.32×10^{-4}), respectively (Fig. 5c). The substitution rate for DENV4 was not available because of the limited number of available DENV4 genomes (Table S4).

Discussion

DF has been prevalent in China over the past three decades, expanding from southeastern coastal provinces to northern China

(Lin et al. 2020). Previous studies have reported the possibility of the local circulation of DENV in parts of China (Zheng et al. 2009, Wu et al. 2011, Zhao et al. 2014, Bai et al. 2018). In countries where DENV is endemic, such as Singapore and Thailand (Lee et al. 2012, Bhoomiboonchoo et al. 2014), DENVs continue to circulate annually, and genomic sequences from different years or locations are characterized by (I) high sequence identity and (II) clustering together on evolutionary trees, as reported in our previous study (Sun et al. 2020). In this study, the DENV genomes from the same or different years or different provinces in China were not clustered together along the ML trees, and most of them were genetically closer to sequences from Southeast Asian countries (Figs. 2 and 3), presenting different characteristics from the countries where DENV is endemic. In addition, frequent tourism and economic exchanges occur between China and Southeast Asian countries. Although the DENV2 genomes from Yunnan and Zhejiang provinces, and the DENV3 genomes from Henan Province, were clustered together (Figs. 2b and 3a), all three dengue outbreaks were strongly associated with imported dengue cases (Huang et al. 2014; Jiang et al. 2018; Yan et al. 2018). This is not sufficient to indicate that the DENV has established a stable circulation in China. Based on the results of this study, we believe that DENV remains an imported pathogen in China.

Genomic recombination was vital for the evolution of DENV, and the diversity of the viral genome was generated to improve its adaptive capacity (Worobey et al. 1999). In this study, we identified 30 intra-serotype recombinant sequences isolated mainly from Guangdong, Yunnan, and Zhejiang provinces (Fig. 4b),

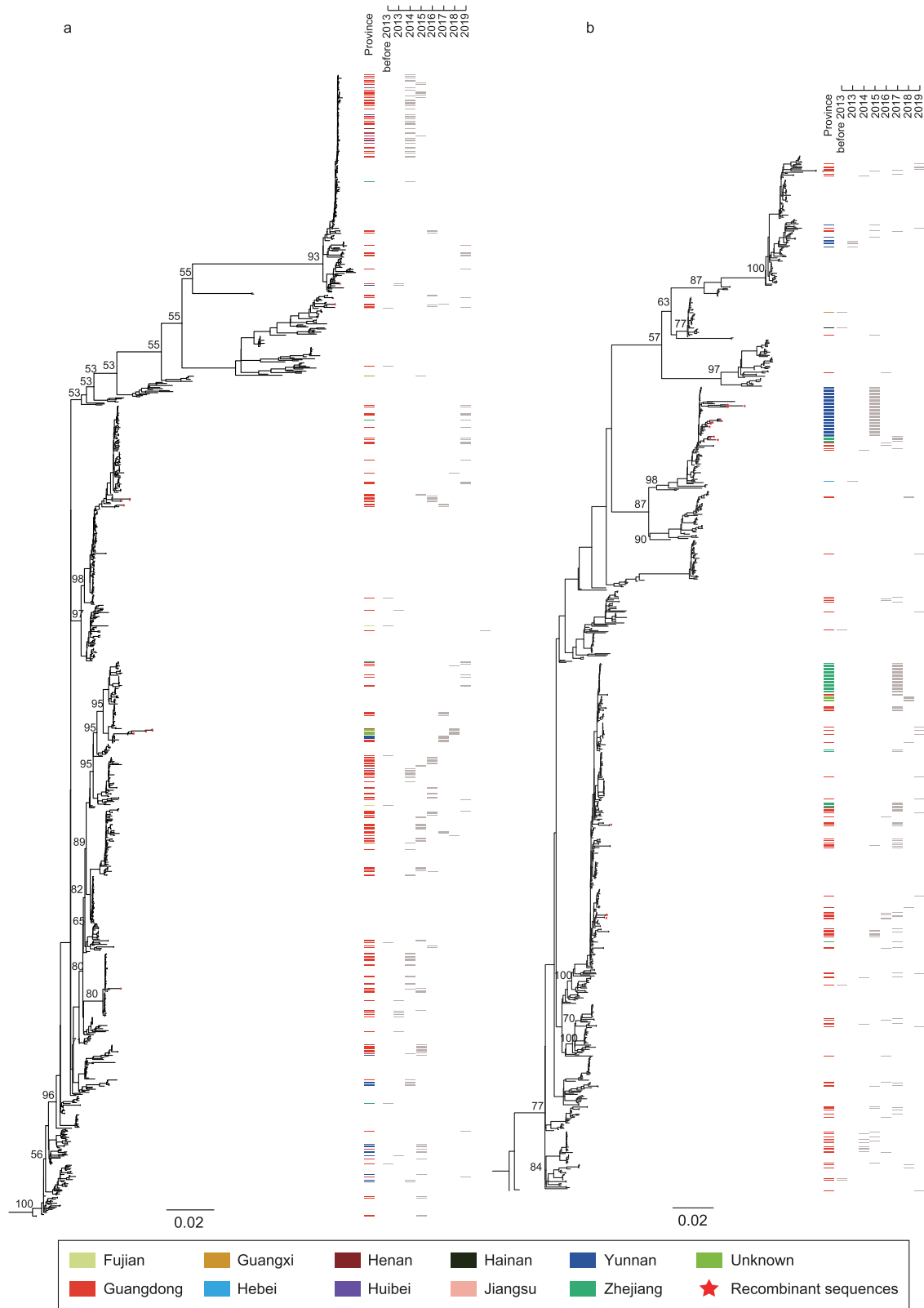


Figure 2. Phylogenetic analysis of DENV1 (a) and DENV2 (b) genomes. The provinces and collection times of DENV genomes in China in this study are represented by colored and gray rectangles, respectively, with the recombinant sequences marked by red stars. Bootstrap values are labeled at major nodes; the scale bar means nucleotide substitutions per site.

some of which have been reported in previous studies, such as KY672959, KY672960, KY937188, and KY937189 from Yunnan (Jiang et al. 2018, Zhu et al. 2022); MH110575 and MH110584

from Zhejiang (Zhu et al. 2022); and MN018340, MN018347, MN018355, MN018372, MN018379, MN018395, MN018396, and MN018397 from Guangdong (Sun et al. 2020, Zhu et al. 2022).

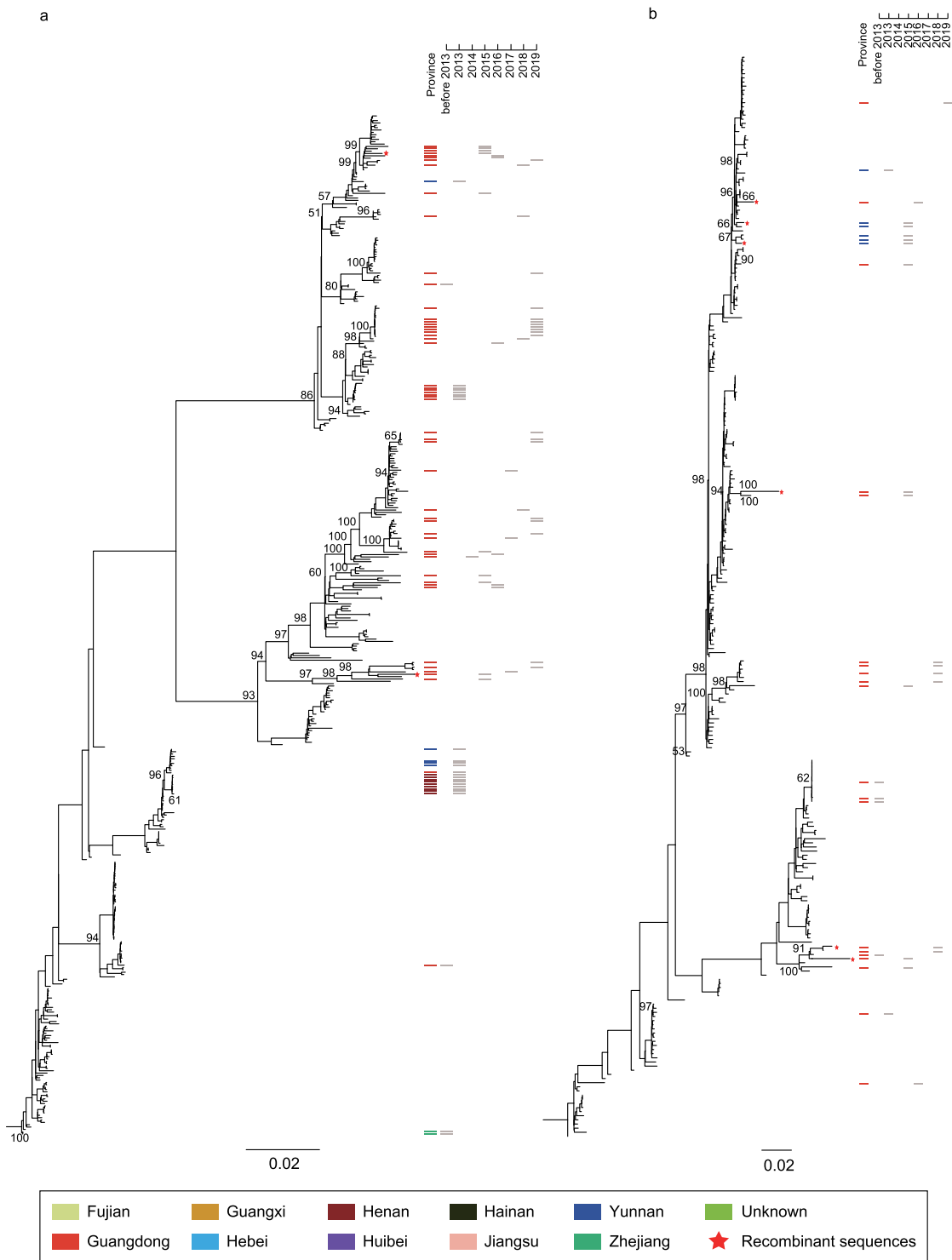


Figure 3. Phylogenetic analysis of DENV3 (a) and DENV4 (b) genomes. The provinces and collection times of DENV genomes in China in this study are represented by colored and gray rectangles, respectively, with the recombinant sequences marked by red stars. Bootstrap values are labeled at major nodes; the scale bar means nucleotide substitutions per site.

However, some other recombinant sequences reported in previous studies were not detected in this study such as KY937186 and MG601754 from Yunnan (Mo et al. 2018, Zhu et al. 2022) and MH827540, MH827543, MN018297, MN018333, MN018334, MN018337, MN018358, MN018359, MN018364, MN018366, MN018378, MN018380, MN018382, MN018393, and MN018394 from Guangdong (Sun et al. 2020, Zhu et al. 2022), in that different

datasets and analytical parameters were used in these recombinant analyses. Phylogenetic analyses revealed that diverse lineages were present in the DENV genomes (Figs. 2 and 3), which was, to some extent, due to frequent viral genomic recombination. The highest frequency of recombination occurred in DENV4 (Fig. 4c), consistent with the results of our previous study on DF in Guangdong Province (Sun et al. 2020). Whether this was

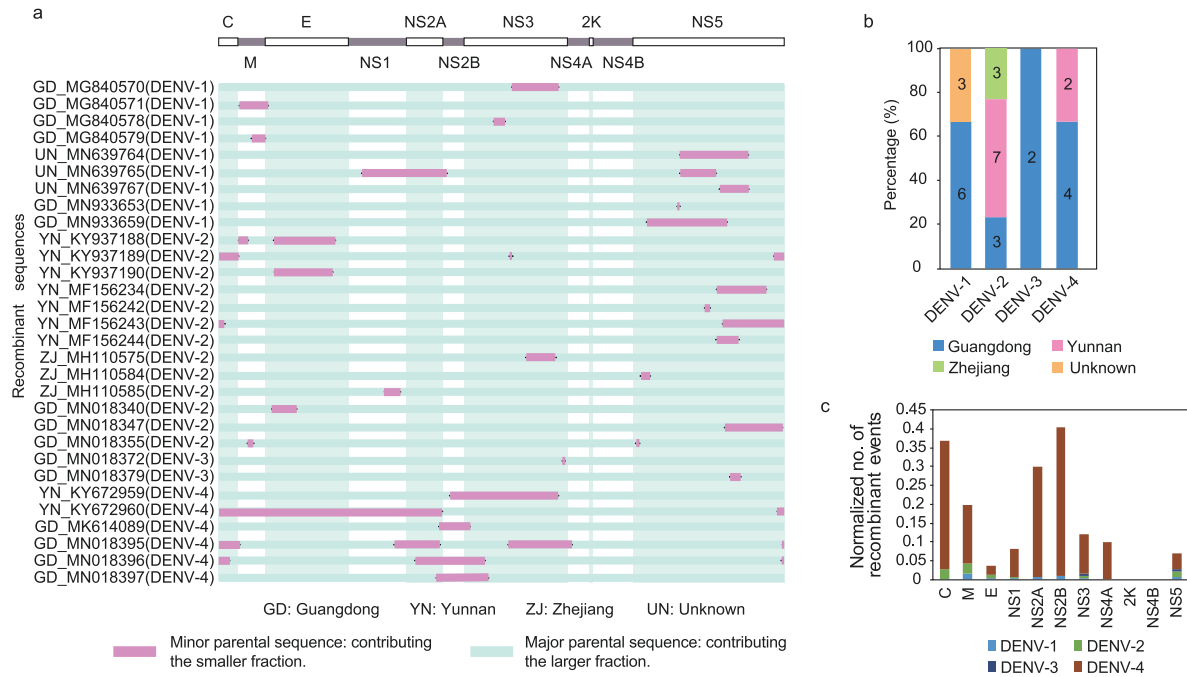


Figure 4. Recombinant detection of DENV genomes in China. (a) Distribution of the recombinant region along the viral genome. Only one DENV genomic structure diagram is represented in the upper panel due to little difference in the entire coding sequence (CDS) regions of DENV1–4. In the low panel, the collection provinces, the GenBank accessions, and the serotypes of the recombinant sequence in this study are marked on the left. (b) Distribution of the number and proportion of DENV1–4 recombinant sequences in the indicated collection provinces in China. (c) Normalized number of recombinant events per gene of DENV genome. Every involved DENV gene was counted one time if the minor parental sequences were across more than one gene. The normalized number of recombinant events was obtained by total recombinant events per DENV gene divided by the sequence number and gene length (kb).

caused by data bias or the biological characteristics of DENV4 requires further investigation. Our results also indicated that (1) some of the recombinant sequences already formed from DENV2 and DENV4 were involved in the generation of new recombinants as parental sequences and (2) some strains acted as parental sequences in the generation of two or more recombinants (Table S2). The generation of recombinant strains indicated that the host was simultaneously infected with at least two genomes and that genetic fragment exchange occurred between them (Pérez-Losada et al. 2015). In this study, the identified recombination sequences and their parental sequences were isolated at different times and locations (Table S2). Therefore, the precise recombination mechanism requires further investigation.

In this study, more synonymous than nonsynonymous substitutions were identified in DENV genomes and most codons of DENV genomes were subject to negative selection pressure, suggesting that purifying selection plays a dominant role in the evolution of DENV in China, consistent with previous reports (Sun et al. 2020, Zhu et al. 2022). Moreover, the distribution characteristics of nonsynonymous substitutions along the viral genome varied among the provinces of China (Fig. 5a). We found some nonsynonymous substitutions that only occurred in specific provinces except Guangdong (DF is so prevalent in Guangdong that we do not believe such substitutions make sense) with a frequency of over 50%, such as G2900A and G3573A of DENV2, T895G of DENV4 in Yunnan, G1291A of DENV2 in Zhejiang, and G6229A of DENV3 in Henan (Table S3); no such substitutions were found in DENV1. These SNPs are the dominant alleles or are fixed in the viral population, but selection pressure analyses showed that these SNPs were under negative selection. Due to the limited number

of Chinese DENV genomes available for this study, the biological significance of these substitutions requires further investigation.

A number of positive selection sites of DENV genomes, which play an important role in the genetic variation of DENV, had been found in previous studies (Bai et al. 2018, Guan et al. 2021, Han et al. 2022, Zhu et al. 2022). In the present study, 13 codon sites under positive selection were identified, suggesting that DENV in China also continues to evolve through positive selection, and thus, future surveillance for DENV needs to be strengthened. Of these 13 codons, only two (codon 3 in DENV1 and codon 9 in DENV2) are located in structural protein regions (the C protein), while the others are located in nonstructural protein regions (Table 2), suggesting that structural proteins are relatively conserved in the evolution of DENV, which is consistent with previous studies (Zhao et al. 2016, Guan et al. 2021). In addition, two positive selection codons (codon 869 and codon 2871 in DENV1) have been reported in previous studies (Han et al. 2022, Zhu et al. 2022), and our investigations provide new insights into understanding the evolution of DENV in China.

The substitution rates of DENV1–4 had been measured based on full-length or partial genomes in previous studies, $\sim 4.60\text{--}11.60 \times 10^{-4}$ s/s/y for different serotypes (Goncalvez et al. 2002, Weaver and Vasilakis 2009, Chen and Vasilakis 2011, Carneiro et al. 2012, Wei and Li 2017, Pollett et al. 2018, Islam et al. 2023). In Guangdong Province, China, the substitution rates of DENV1–4 were $\sim 7.73 \times 10^{-4}$ s/s/y (95% HPD, 7.00×10^{-4} to 8.43×10^{-4}), 8.34×10^{-4} s/s/y (95% HPD, 7.67×10^{-4} to 9.00×10^{-4}), 9.33×10^{-4} s/s/y (95% HPD, 8.03×10^{-4} to 1.07×10^{-3}), and 1.10×10^{-3} s/s/y (95% HPD, 9.48×10^{-4} to 1.25×10^{-3}), respectively, as calculated based on the E gene in a previous study (Cui et al. 2022).

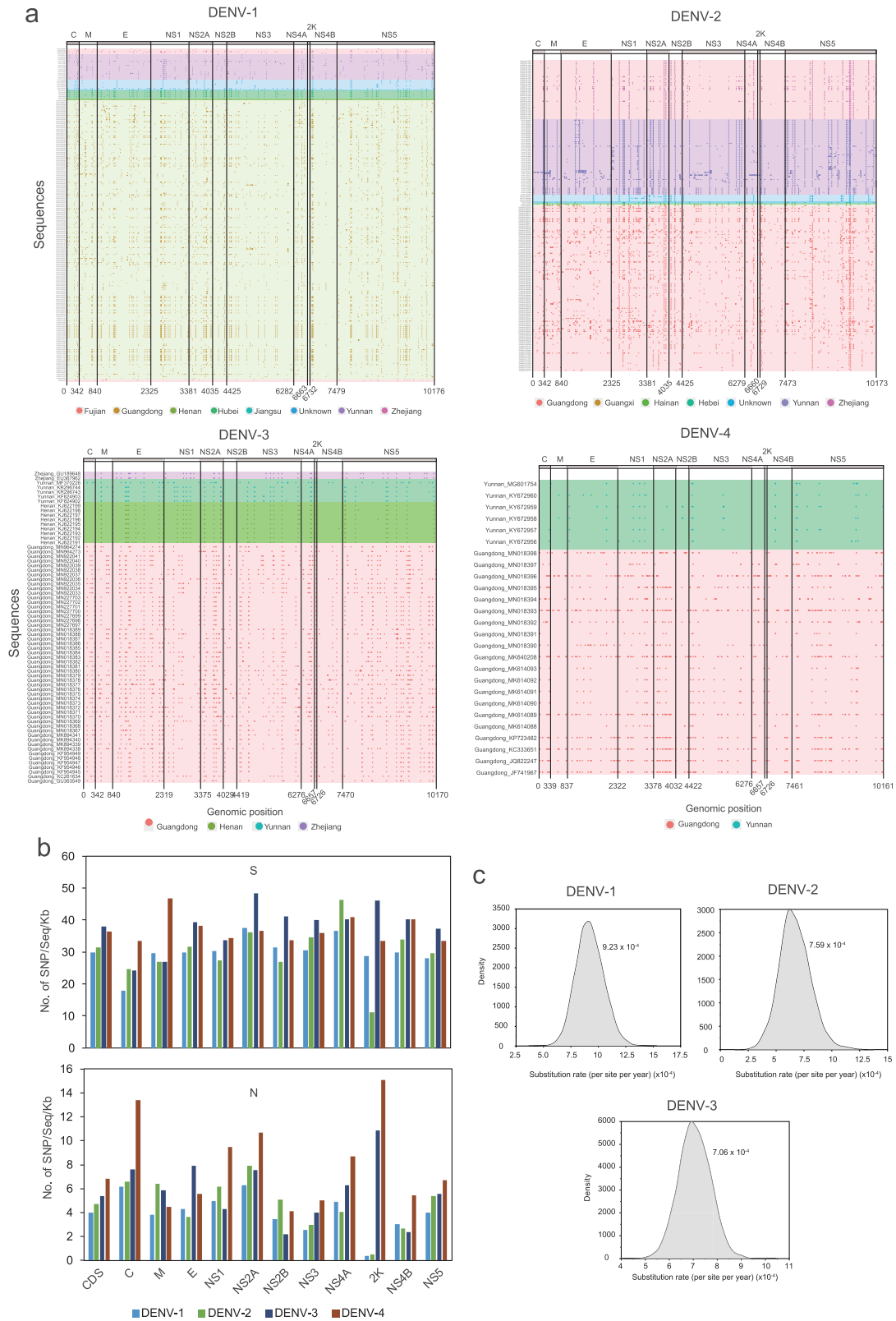


Figure 5. The substitutions of DENV genomes in China. (a) Distribution of the nonsynonymous substitutions along the DENV genome in indicated serotypes. The nonsynonymous substitutions from different provinces of China are marked with colored dots; the provinces and GenBank accessions are marked on the left. (b) The normalized SNP among each gene of DENV1–4 genomes. The normalized SNP per gene of DENV genomes was obtained by total SNPs per DENV gene divided by the sequence number and gene length (kb) and was compared with that of CDS using all pairwise Kruskal–Wallis one-way Analysis of Variance (ANOVA) tests with $P < .01$. N, nonsynonymous; S, synonymous. (c) The nucleotide substitution rates of DENV1–3 genomes in China.

Table 2. The codon sites of DENV1–4 under positive selection.

Serotype	Codon	Gene	FEL		MEME		SLAC		FUBAR	
			dN/dS	P	dN/dS	P	dN/dS	P	dN/dS	Prob ($\alpha < \beta$)
1	3	C	∞	.081	∞	.000	1.995	.053	2.013	.973
1	869	NS1	6.882	.017	29.200	.010	2.242	.066	1.913	.960
1	1490	NS3	∞	.038	∞	.050	1.807	.077	–	–
1	2628	NS5	∞	.027	∞	.040	3.933	.048	–	–
1	2871	NS5	6.820	.015	6.864	.020	2.808	.033	–	–
1	3133	NS5	2.892	.058	2.871	.080	–	–	4.940	.904
2	9	C	∞	.040	∞	.060	1.643	.078	–	–
2	1298	NS2A	∞	.032	∞	.050	3.435	.031	–	–
2	1867	NS3	5.438	.040	10.985	.050	2.116	.055	–	–
3	1112	NS1	∞	.019	∞	.030	1.881	.097	–	–
4	927	NS1	4.746	.058	12.321	.020	2.509	.055	–	–
4	1145	NS2A	∞	.023	∞	.030	1.739	.055	–	–
4	3118	NS5	5.542	.046	5.444	.060	–	–	6.253	.934

Other studies reported that the average substitution rate of DENV1 in Guangdong was 1.03×10^{-3} s/s/y (95% HPD, 7.36×10^{-4} to 1.34×10^{-3}) (Bai et al. 2018) and that of DENV2 in Zhejiang was 3.08×10^{-3} s/s/y (95% HPD, 1.56×10^{-3} to 4.64×10^{-3}) (Yu et al. 2019). In this study, the DENV1–3 substitution rates evaluated using genome-wide data were 9.23×10^{-4} (95% HPD, 6.86×10^{-4} to 1.17×10^{-3}), 7.59×10^{-4} (95% HPD, 3.90×10^{-4} to 9.89×10^{-4}), and 7.06×10^{-4} s/s/y (95% HPD, 5.76×10^{-4} to 8.32×10^{-4}), respectively (Fig. 5c). The DENV substitution rates from previous studies (Bai et al. 2018, Cui et al. 2022) were calculated based on the E gene, whereas our results were based on the whole genome of DENV. In addition, our investigation differs from that in the study by Yu et al. (2019) in terms of dataset, computational model, chain length, etc., which explains the inconsistency in the above substitution rates. In particular, the differences in the collection time of the DENV genome also affected the substitution rates. Generally, the substitution rate in the long term (between disease outbreaks) is lower than that in the short term (within disease outbreaks), because many deleterious variants are eliminated by purifying selection (Holmes et al. 2016). The genomic data in this study were mainly collected from 2012 to 2019 and covered only 10 provinces (Table 1 and Fig. 1). Hence, the evolutionary rate of DENV in China requires further investigation.

In summary, we elucidated the phylogenetics, recombination, selection pressure, and nucleotide variants of DENV in China and assessed their substitution rates based on whole viral genomes, offering further knowledge of DENV evolution in China. However, the unavailability of whole genomes in provinces with dengue epidemics, combined with missing genomes owing to the coronavirus disease (COVID-19) pandemic from 2020 to 2022, limits our study. With the end of the global COVID-19 pandemic, increasingly frequent exchanges between China and other countries have posed additional challenges for dengue prevention and control. Our investigation indicates that recombination and variants of DENV genomes are frequent, although DENV is not yet endemic in China. Therefore, dengue surveillance needs to be strengthened in the future.

Supplementary data

Supplementary data is available at *VEVOLU Journal* online.

Conflict of interest: None declared.

Funding

This work was supported by the Natural Science Foundation of Shandong Province of China (ZR2022QC202, ZR2023QH354, and ZR2022QH126), Natural Science Foundation of Hubei Province of China (2022CFB837, 2023AFB221, and 2021CFA012), and National Natural Science Foundation of China (31970548).

Data availability

Data are available on request.

References

- Bai Z, Liu L, Jiang L et al. Evolutionary and phylodynamic analyses of dengue virus serotype I in Guangdong Province, China, between 1985 and 2015. *Virus Res* 2018;**256**:201–08.
- Bartolucci F, Bellio R, Salvan A et al. Modified profile likelihood for fixed-effects panel data models. *Econometric Rev* 2016;**35**: 1271–89.
- Bhoomiboonchoo P, Gibbons R, Huang A et al. The spatial dynamics of dengue virus in Kamphaeng Phet, Thailand. *PLoS Negl Trop Dis* 2014;**8**:e3138.
- Boni MF, Posada D, Feldman MW. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* 2007;**176**:1035–47.
- Carneiro A, Cruz A, Vallinoto M et al. Molecular characterisation of dengue virus type 1 reveals lineage replacement during circulation in Brazilian territory. *Memorias Do Instituto Oswaldo Cruz* 2012;**107**:805–12.
- Chen R, Vasilakis N. Dengue—quo tu et quo vadis?. *Viruses* 2011;**3**:1562–608.
- Cui F, He F, Huang X et al. Dengue and dengue virus in Guangdong, China, 1978–2017: epidemiology, seroprevalence, evolution, and policies. *Front Med Lausanne* 2022;**9**:797674.
- Du J, Zhang L, Hu X et al. Phylogenetic analysis of the dengue virus strains causing the 2019 dengue fever outbreak in Hainan, China. *Virologica Sin* 2021;**36**:636–43.
- Gibbs MJ, Armstrong JS, Gibbs AJ. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 2000;**16**:573–82.
- Goncalvez A, Escalante A, Pujol F et al. Diversity and evolution of the envelope gene of dengue virus type 1. *Virology* 2002;**303**: 110–19.

- Guan J, Li Z, Chen J et al. Complete genome characterization of the 2018 dengue outbreak in Hunan, an inland province in central South China. *Virus Res* 2021;**297**:198358.
- Han A, Sun B, Sun Z et al. Molecular characterization and phylogenetic analysis of the 2019 Dengue Outbreak in Wenzhou, China. *Front Cell Infect Microbiol* 2022;**12**:829380.
- Holmes E, Dudas G, Rambaut A et al. The evolution of Ebola virus: insights from the 2013–2016 epidemic. *Nature* 2016;**538**:193–200.
- Huang X, Ma H, Wang H et al. Outbreak of dengue fever in central China, 2013. *Biomed Environ Sci* 2014;**27**:894–97.
- Islam A, Deeba F, Tarai B et al. Global and local evolutionary dynamics of dengue virus serotypes 1, 3, and 4. *Epidemiol Infect* 2023;**151**:e127.
- Jiang L, Ma D, Ye C et al. Molecular characterization of dengue virus serotype 2 cosmopolitan genotype from 2015 dengue outbreak in Yunnan, China. *Front Cell Infect Microbiol* 2018;**8**:219.
- Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evolu* 2005;**22**:1208–22.
- Kyle JL, Harris E. Global spread and persistence of dengue. *Annu Rev Microbiol* 2008;**62**:71–92.
- Lee K, Lo S, Tan S et al. Dengue virus surveillance in Singapore reveals high viral diversity through multiple introductions and in situ evolution. *Infect Genet Evol* 2012;**12**:77–85.
- Lin H, Wang X, Li Z et al. Epidemiological characteristics of dengue in mainland China from 1990 to 2019: a descriptive analysis. *Medicine* 2020;**99**:e21982.
- Liu Q, Wang J, Hou J et al. Entomological investigation and detection of dengue virus type 1 in *Aedes (Stegomyia) albopictus* (Skuse) during the 2018–2020 outbreak in Zhejiang Province, China. *Front Cell Infect Microbiol* 2022;**12**:834766.
- Ma M, Wu S, He Z et al. New genotype invasion of dengue virus serotype 1 drove massive outbreak in Guangzhou, China. *Parasites Vectors* 2021;**14**:126.
- Martin DP, Murrell B, Golden M et al. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evolu* 2015;**1**:vev003.
- Martin D, Posada D, Crandall K et al. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retroviruses* 2005;**21**:98–102.
- Meng J, Hu Q, Zhang L et al. Isolation and genetic evolution of dengue virus from the 2019 outbreak in Xishuangbanna, Yunnan Province, China. *Vector Borne Zoonotic Dis* 2023;**23**:331–40.
- Mo L, Shi J, Guo X et al. Molecular characterization of an imported dengue virus serotype 4 isolate from Thailand. *Arch Virol* 2018;**163**:2903–06.
- Murrell B, Moola S, Mabona A et al. FUBAR: a fast, unconstrained Bayesian approximation for inferring selection. *Mol Biol Evolu* 2013;**30**:1196–205.
- Murrell B, Wertheim JO, Moola S et al. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 2012;**8**:e1002764.
- Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent recombination. *Virology* 1999;**265**:218–25.
- Pérez-Losada M, Arenas M, Galán JC et al. Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. *Infect Genet Evol* 2015;**30**:296–307.
- Pollett S, Melendrez M, Maljkovic Berry I et al. Understanding dengue virus evolution to support epidemic surveillance and counter-measure development. *Infect Genet Evol* 2018;**62**:279–95.
- Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci USA* 2001;**98**:13757–62.
- Remco B, Joseph H, Denise K et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Computat Biol* 2014;**10**:e1003537.
- Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends genetics* 2000;**16**:276–77.
- Sang S, Chen B, Wu H et al. Dengue is still an imported disease in China: a case study in Guangzhou. *Infect Genet Evol* 2015;**32**:178–90.
- Sang S, Liu Q, Guo X et al. The epidemiological characteristics of dengue in high-risk areas of China, 2013–2016. *PLoS Negl Trop Dis* 2021;**15**:e0009970.
- Shepard D, Undurraga E, Halasa Y et al. The global economic burden of dengue: a systematic analysis. *Lancet Infect Dis* 2016;**16**:935–41.
- Smith JM. Analyzing the mosaic structure of genes. *J Mol Evolu* 1992;**34**:126–29.
- Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;**30**:1312–13.
- Suchard M, Lemey P, Baele G et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolu* 2018;**4**:vey016.
- Sun B, Zhang X, Zhang H et al. Genomic epidemiological characteristics of dengue fever in Guangdong province, China from 2013 to 2017. *PLoS Negl Trop Dis* 2020;**14**:e0008049.
- Sun H, Yao W, Siddique A et al. Genomic characterization of dengue virus serotype 2 during dengue outbreak and endemics in Hangzhou, Zhejiang (2017–2019). *Front Microbiol* 2023;**14**:1245416.
- Tayal A, Kabra S, Lodha R. Management of dengue: an updated review. *Indian J Pediatr* 2023;**90**:168–77.
- Wang R, Fan D, Wang L et al. Seroprevalence of dengue virus among young adults in Beijing, China, 2019. *Virologica Sin* 2021;**36**:333–36.
- Weaver S, Vasilakis N. Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. *Infect Genet Evol* 2009;**9**:523–40.
- Wei K, Li Y. Global evolutionary history and spatio-temporal dynamics of dengue virus type 2. *Sci Rep* 2017;**7**:45505.
- WHO. 2021; <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>
- Wilder-Smith A, Ooi E-E, Horstick O et al. Dengue. *Lancet* 2019;**393**:350–63.
- Worobey M, Rambaut A, Holmes E. Widespread intra-serotype recombination in natural populations of dengue virus. *Proc Natl Acad Sci USA* 1999;**96**:7352–57.
- Wu T, Wu Z, Li YP. Dengue fever and dengue virus in the People's Republic of China. *Rev Med Virol* 2022;**32**:e2245.
- Wu W, Bai Z, Zhou H et al. Molecular epidemiology of dengue viruses in southern China from 1978 to 2006. *Virol J* 2011;**8**:322.
- Xu L, Stige L, Chan K et al. Climate variation drives dengue dynamics. *Proc Natl Acad Sci USA* 2017;**114**:113–18.
- Xu M, Chang N, Tu T et al. Economic burden of dengue fever in China: a retrospective research study. *PLoS Negl Trop Dis* 2022;**16**:e0010360.
- Yan H, Ding Z, Yan J et al. Epidemiological characterization of the 2017 dengue outbreak in Zhejiang, China and molecular characterization of the viruses. *Front Cell Infect Microbiol* 2018;**8**:216.
- Yao W, Yang Z, Lou X et al. Molecular characterization of dengue virus type 1 in Zhejiang in 2019. *Front Cell Infect Microbiol* 2021;**11**:673299.
- Yu H, Kong Q, Wang J et al. Multiple lineages of dengue virus serotype 2 cosmopolitan genotype caused a local dengue outbreak in Hangzhou, Zhejiang Province, China, in 2017. *Sci Rep* 2019;**9**:7345.

- Yue Y, Liu Q, Liu X *et al.* Comparative analyses on epidemiological characteristics of dengue fever in Guangdong and Yunnan, China, 2004-2018. *BMC Public Health* 2021;**21**: 1389.
- Yue Y, Liu X, Xu M *et al.* Epidemiological dynamics of dengue fever in mainland China, 2014-2018. *Int J Infect Dis* 2019;**86**: 82-93.
- Zh W, Tong C, Decheng W *et al.* Genomic surveillance and evolutionary dynamics of type 2 porcine reproductive and respiratory syndrome virus in China spanning the African swine fever outbreak. *Virus Evol* 2024;**10**: veae016.
- Zhang J, Shu Y, Shan X *et al.* Co-circulation of three dengue virus serotypes led to a severe dengue outbreak in Xishuangbanna, a border area of China, Myanmar, and Laos. 2019 *Int J Infect Dis* 2021;**107**:15-17.
- Zhang Y, Wang L, Wang G *et al.* An ecological assessment of the potential pandemic threat of dengue virus in Zhejiang province of China. *BMC Infect Dis* 2023;**23**:473.
- Zhao H, Zhao L, Jiang T *et al.* Isolation and characterization of dengue virus serotype 2 from the large dengue outbreak in Guangdong, China in 2014. *Sci China Life Sci* 2014;**57**:1149-55.
- Zhao Y, Li L, Ma D *et al.* Molecular characterization and viral origin of the 2015 dengue outbreak in Xishuangbanna, Yunnan, China. *Sci Rep* 2016;**6**:34444.
- Zheng K, Zhou HQ, Yan J *et al.* Molecular characterization of the E gene of dengue virus type 1 isolated in Guangdong province, China, 2006. *Epidemiol Infect* 2009;**137**:73-78.
- Zhu X, Chen W, Ma C *et al.* Whole genome analysis identifies intra-serotype recombinants and positive selection sites of dengue virus in mainland China from 2015 to 2020. *Virus Res* 2022;**311**:198705.