

RESEARCH ARTICLE

# The evolutionary dynamics of DENV 4 genotype I over a 60-year period

Shaowei Sang<sup>1,2\*</sup>, Jing Liu-Helmerson<sup>3</sup>, Mikkel B. M. Quam<sup>3</sup>, Hongning Zhou<sup>4</sup>, Xiaofang Guo<sup>4</sup>, Haixia Wu<sup>5</sup>, Qiyong Liu<sup>5\*</sup>

**1** Clinical Epidemiology Unit, Qilu Hospital of Shandong University, Jinan, Shandong, People's Republic of China, **2** Department of Epidemiology and Health Statistics, School of Public Health, Shandong University, Jinan, Shandong, People's Republic of China, **3** Department of Epidemiology and Global Health, Umea University, Umea, Sweden, **4** Yunnan Provincial Center of Arbovirus Research, Yunnan Provincial Key Laboratory of Vector-borne Diseases Control and Research, Yunnan Institute of Parasitic Diseases, Pu'er, Yunnan, People's Republic of China, **5** State Key Laboratory of Infectious Disease Prevention and Control, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Changping, Beijing, People's Republic of China

\* [sangshaowei@sdu.edu.cn](mailto:sangshaowei@sdu.edu.cn) (SS); [liuqiyong@icdc.cn](mailto:liuqiyong@icdc.cn) (QL)



**OPEN ACCESS**

**Citation:** Sang S, Liu-Helmerson J, Quam MBM, Zhou H, Guo X, Wu H, et al. (2019) The evolutionary dynamics of DENV 4 genotype I over a 60-year period. *PLoS Negl Trop Dis* 13(7): e0007592. <https://doi.org/10.1371/journal.pntd.0007592>

**Editor:** Hannah Clapham, University of Oxford, VIET NAM

**Received:** November 14, 2018

**Accepted:** July 1, 2019

**Published:** July 29, 2019

**Copyright:** © 2019 Sang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its Supporting Information files.

**Funding:** SS was supported by the National Natural Science Foundation of China (NSFC) (Grant No.81602898) and China Postdoctoral Science Foundation (2018M630790). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Dengue virus serotype 4 (DENV 4) has had a relatively low prevalence worldwide for decades; however, likely due to data paucity, no study has investigated the epidemiology and evolutionary dynamics of DENV 4 genotype I (DENV 4-I). This study aims to understand the diversity, epidemiology and dynamics of DENV 4-I. We collected 404 full length DENV4-1 envelope (E) gene sequences from 14 countries using two sources: Yunnan Province in China (15 strains during 2013–2016) and GenBank (489 strains up to 2018-01-11). Conducting phylogenetic and phylogeographical analyses, we estimated the virus spread, population dynamics, and selection pressures using different statistical analysis methods (substitution saturation, likelihood mapping, Bayesian coalescent inference, and maximum likelihood estimation). Our results show that during the last 60 years (1956–2016), DENV 4-I was present in mainland and maritime Southeast Asia, the Indian subcontinent, the southern provinces of China, parts of Brazil and Australia. The recent spread of DENV 4-I likely originated in the Philippines and later spread to Thailand. From Thailand, it spread to adjacent countries and eventually the Indian subcontinent. Apparently diverging around years 1957, 1963, 1976 and 1990, the different Clades (Clade I-V) were defined. The mean overall evolution rate of DENV 4-I was  $9.74$  (95% HPD:  $8.68$ – $10.82$ )  $\times 10^{-4}$  nucleotide substitutions/site/year. The most recent common ancestor for DENV 4-I traces back to 1956. While the demographic history of DENV 4-I fluctuated, peaks appeared around 1982 and 2006. While purifying selection dominated the majority of E-gene evolution of DENV 4-I, positive selection characterized Clade III (Vietnam). DENV 4-I evolved in situ in Southeast Asia and the Indian subcontinent. Thailand and Indian acted as the main and secondary virus distribution hubs globally and regionally. Our phylogenetic analysis highlights the need for strengthened regional cooperation on surveillance and sharing of sample sequences to improve global dengue control and cross-border transmission prevention efforts.

**Competing interests:** The authors have declared that no competing interests exist.

## Author summary

Dengue virus (DENV) can be classified into four serotypes, DENV 1, 2, 3 and 4. Although DENV 4 is the first dengue serotype to diverge in phylogenetic analyses of the genus *Flavivirus*, this serotype occurs at a low prevalence worldwide and spreads the least rapidly. Similar to other serotypes, DENV 4 can also cause severe dengue (SD) disease manifestations, such as dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS). To date, no study has investigated the epidemiology and dynamics of DENV 4 genotype I comprehensively. In this study, we seek to address this gap. Our study shows that the distribution of DENV 4-I is mainly restricted to Southeast Asia and the Indian subcontinent. The most recent spread of DENV 4-I likely originated from Southeast Asia—initially circulating in the Philippines, then Thailand and later on the Indian subcontinent. Viruses evolved in situ in Southeast Asia and the Indian subcontinent, respectively. Although DENV 4-I occasionally spread elsewhere, this genotype did not become widely established. The overall evolution rate of DENV 4-I was comparable with that of DENV 2–4. The nucleotide sequences indicates that the demographic history of DENV 4-I fluctuated with peaks apparent during parts of the 1980s and 2000s. Although a weak positive selection existed in Clade III -predominately in Vietnam, purifying selection dominated the E-gene evolution of DENV 4-I.

## Introduction

Dengue is a mosquito-borne viral infectious disease. Although the geographical origin of dengue is still under some debate, the recent global expansion has been attributed to environmental changes, unprecedented population growth, uncontrolled urbanization, spread of the mosquito vectors, and host population movements [1]. Currently, dengue is endemic in more than 100 countries in much of the globe's tropical and subtropical areas, being reported predominantly in Southeast Asia, the Americas, and the Western Pacific, and less frequently in Africa and Eastern Mediterranean WHO regions [2]. The prevalence of dengue has increased 74.7% between 2006 and 2016 [3]. While dengue infections are most often asymptomatic, a recent study has estimated a global total of 58.4 (95% CI: 24–122) million symptomatic dengue cases occur annually costing some US\$8.9 (95% CI: 3.7–19.7) billion [4].

The four antigenically distant serotypes comprising dengue virus are categorized as DENV 1, -2, -3, and -4. Each serotype is classified into different genotypes based on complete E gene sequences [5]. The four serotypes were first identified at different times and locations, and diffused globally at different rates. Although clearly circulating before isolation techniques enabled the viruses' discovery and characterization, DENV 1 was first reported in French Polynesia and Japan, DENV 2 in 1944 in Papua New Guinea and Indonesia, DENV 3 and DENV 4 both in 1953 in the Philippines and Thailand [6]. A study mapping the global spread of DENV 1–4 over the 70-year history (1943–2013), indicated DENV 1 was reported most often, followed by DENV 2, DENV 3, and DENV 4 [7]. Although, DENV 4 was the first dengue serotype to diverge in phylogenetic analyses of the genus *Flavivirus* [8], it spread the slowest geographically [7]. Similar to other serotypes, DENV 4 can also cause severe dengue including dengue haemorrhagic fever (DHF). An epidemiological study in Thailand from 1973 to 1999 revealed that despite the proportionately lower prevalence of DENV 4, it was responsible for 10% of all DHF cases in Children [9].

Among DENV 4, there are four genotypes (I, II, III and Sylvatic). The dengue cases from DENV 4 genotype I (DENV 4-I) have increased in recent years. In 2013, a large DENV

outbreak occurred in central Vietnam with a total of 204,661 clinical cases reported, of which 48.9% were DENV 4-I cases [10]. The same year, DENV 1-I, DENV 2-I and DENV 4-I caused a large outbreak with 20,255 cases including 84 deaths in Myanmar [11]. In Sri Lanka, the dominant virus in the 2012 epidemic was DENV 1, but DENV 4-I infections were also commonly observed [12]. While all four serotypes have been detected, a 2015 study showed that dengue in China remains primarily an imported disease with DENV 1 most frequently found in samples [13]. However, since 2013 several strains of DENV 4-I have been detected in Yunnan Province, China. Among the recorded DENV 4-I strains, both imported and autochthonous cases were found.

The dengue viruses DENV 1–4 are typically prevalent in tropical and subtropical regions globally. However, the spatial distribution of different genotypes is not uniform, e.g., some genotypes exist only in specific parts of Asia and others are more cosmopolitan. While the distinct distribution patterns of different genotypes remain enigmatic, mapping of the genotypes' distribution can generate hypotheses on their spatial pattern and support policies on dengue prevention and control effort. In the past, efforts have been made to infer the dispersal of DENV 1–4 and to elucidate the evolution of their diffusion patterns [14–17]. However, so far no studies investigated globally the spatial distribution of the single genotype, DENV 4-I, its diversity, and its temporal evolution. This may be due to the limited number of recorded cases and the distribution of DENV 4-I worldwide.

In this study, we used available GenBank data in addition to Chinese data sources to make the first attempt to more comprehensively understand the spatial and evolutionary patterns of DENV 4-I. Leveraging full envelope gene sequences in our analysis, we sought to investigate the origin and spreading routes of this less-studied, rare but deadly virus, in order to contribute important information for future dengue prevention and control efforts around the globe.

## Methods

### Ethics statement

Ethical approval for the study was obtained from the Chinese Center for Disease Control and Prevention Ethical Committee (No. 201214).

### Data collection

Dengue viruses detected in Yunnan Province were recovered from serum samples of suspected dengue patients visiting hospital from 2013 to 2016. The envelope (E) genes of isolates were sequenced as described previously [18]. These have been assigned GenBank accession numbers (HM893690–HM893699, MG601754, KJ470764, KJ470765, KX262920, KX262923). The sequences of Yunnan were compared with published sequences by using the nucleotide blast program in the NCBI. All the sequences of human DENV 4-I with full length E-gene (1,485 nucleotides) were downloaded with the accession number, collection date and country/region (as of January 11<sup>th</sup>, 2018). All the sequences were aligned using MAFFT [19]. We chose only one sequence to represent sequences with 100% matching identity by location and time. Recombinants detected based on the analyses of RDP3 program [20] were also excluded. Ultimately, 404 sequences of DENV 4-I obtained from 14 countries were included for analyses in this study (S1 Table).

### Substitution saturation analysis and likelihood mapping analysis

The phylogenetic signal of the aligned DENV 4-I was evaluated by plotting the observed number of transitions and transversions against genetic distance for the  $n(n-1)/2$  pairwise comparisons in an alignment of  $n$  taxa using DAMBE [21]. It is expected that transitions and

transversions increase linearly with the genetic distance, with transitions being more frequent than transversions. In the Supplementary Materials, [S1 Fig](#) shows that no substitution saturation was detected, indicating phylogeny reconstruction was appropriate. In likelihood mapping analysis, groups of four sequences (quartets) were evaluated using the maximum likelihood approach. For each quartet, the three possible unrooted tree topologies were weighted. The likelihood weights were then plotted into a triangular surface. The fully resolved tree topologies were plotted in the three corners, which indicated the presence of a tree-like phylogenetic signal, and the unresolved quartets, indicating a star-like signal were shown in the central region of the triangle [22]. Likelihood mapping was performed using the TREE-PUZZLE program [23], by analyzing 10,000 random quartets. [S1 Fig](#) showed tree-like area accounted for 96.4%, which further suggested that the data were reliable for phylogenetic inference.

### Phylogenetic and phylogeographical analyses

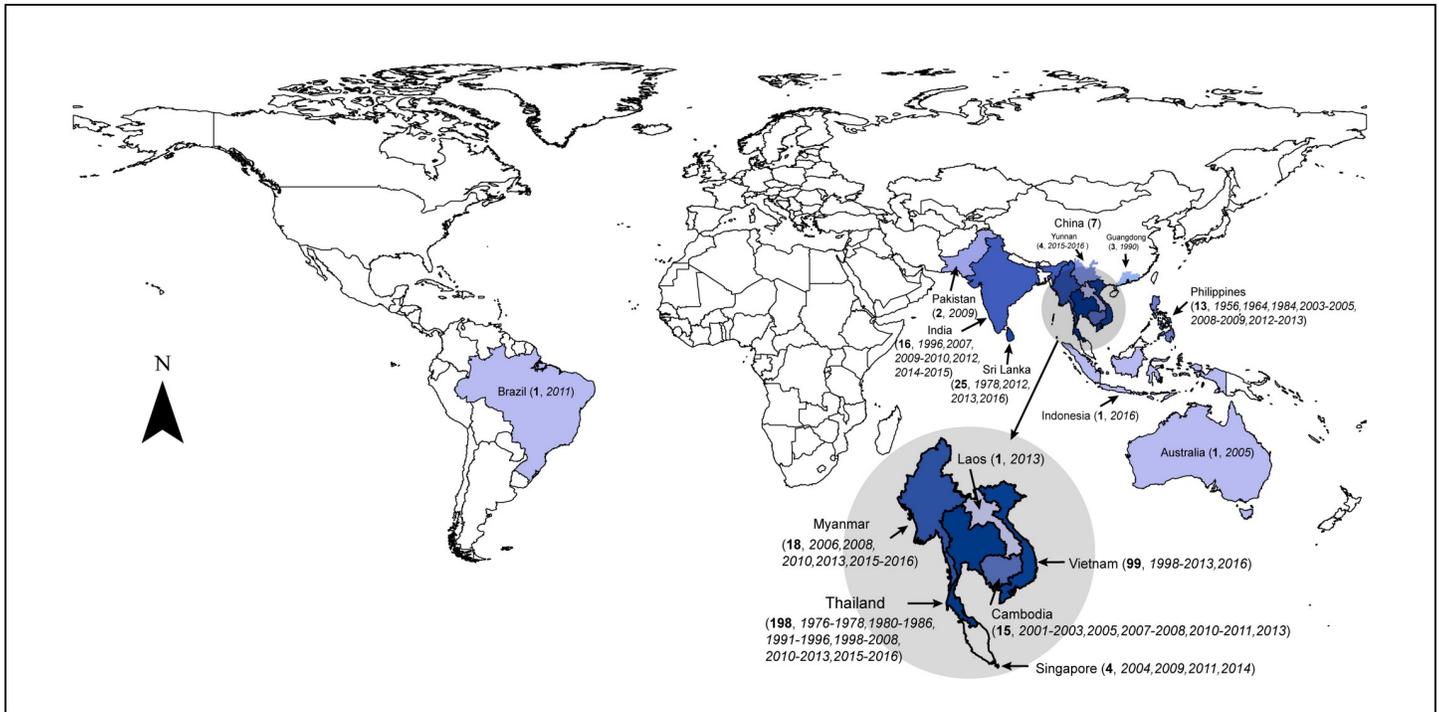
Rates of nucleotide substitution per site per year and time to The Most Recent Common Ancestor (TMRCA) were estimated using Bayesian Markov Chain Monte Carlo (MCMC) and implemented using the BEAST v1.8.2 software package [24]. The best-fit model of nucleotide substitutions was determined using Bayesian Information Criteria (BIC) as implemented in jModelTest [25]. The calibration point was the “year” that each strain was isolated. Statistical simulations were performed under strict or relaxed (uncorrelated exponential and lognormal) clock model, with the Bayesian Skyride Coalescent Tree Prior [26]. To determine the best-fit combination, we have applied Posterior-simulation Akaike Information Criterion through MCMC (AICM) [27], Bayes Factor (BF) [28], Harmonic Mean (HM) [29], and Path Sampling (PS) and Stepping-stone Sampling (SS) [30] model selection methods. The results showed that the best fitting model was the combination of uncorrelated relaxed exponential clock model and the Bayesian Skyride Coalescent model ([S2 Table](#)). Statistical uncertainties in parameter values were given by the 95% Highest Probability Density (HPD) intervals. All chains were run sufficiently long to achieve convergence (the effective sample size of continuous parameters greater than 200) after burn-in, as checked using TRACER v1.5 (<http://tree.bio.ed.ac.uk/software/tracer/>). The programs TreeAnnotator v1.8.2 in the BEAST v1.8.2 software package and Figtree (<http://tree.bio.ed.ac.uk/software/figtree/>) were used to summarize the posterior tree distribution and to visualize the annotated Maximum Clade Credibility (MCC) tree, respectively. Based on the MCC tree, we identified five Clades using visual judgement and comparison among all the countries/regions that reported DENV 4-I. Using the definition of a minimum of three sequences of monophyletic origin, DENV 4-I were labelled with Clade I to V (the largest five) where every Clade included as many strains as possible.

The spatial diffusion of DENV 4-I was estimated using the Bayesian Markov chain Monte Carlo (MCMC) statistical framework implemented in the BEAST v1.8.2 package. The phylogeographical diffusion process was identified using the Bayesian Stochastic Variable Search Selection (BSVSS). Effective population size dynamics were estimated using the Bayesian Skyride Coalescent statistical approach.

Open source data from <http://tapiquen-sig.jimdo.com> (Carlos Efraín Porto Tapiquén. Oro-génesis Soluciones Geográficas. Porlamar, Venezuela, 2015) were used in this study for the results shown in [Figs 1 & 4](#) with help of ArcGIS 10.2 and Adobe Illustrator.

### Estimating selection pressure

We used a variety of computational methods to explore the selection pressures. A Maximum Likelihood (ML) method was used to examine selection pressures [31]. In the analysis, the non-synonymous to synonymous rate ratio ( $\omega = d_N/d_S$ ) was determined codon-by-codon



**Fig 1. The global distribution of DENV 4-I.** The number of unique strains (in bold, < 100% similarity in the same year) and the years of recorded transmission (in italic) in each location are shown in parenthesis after each location name.

<https://doi.org/10.1371/journal.pntd.0007592.g001>

using various models of codon substitution. These models differ in how  $\omega$  ratios are allowed to vary along the sequence. Four models of codon substitution were conducted in the study: M1a ( $\omega \leq 1$ ; nearly neutral), M2a ( $\omega \leq 1$  and  $\omega > 1$ ; positive selection), M7 (beta; a discrete distribution with 10 site classes to model values of  $\omega$  between 0 and 1) and M8 (beta and  $\omega > 1$ ). M1a is nested with M2a, and M7 is nested with M8. Models that are nested are compared statistically using a Likelihood Ratio Test (LRT). Positive selection can be inferred when a group of codons with a  $\omega$  ratio  $> 1$  is identified and the likelihood of the codon substitution model in question is significantly higher ( $p < 0.05$ ) than that of a nested model that does not take positive selection into account. Lastly, using Bayes Empirical Bayes (BEB) methods, posterior probabilities were calculated to identify sites under positive selection (posterior Bayesian probability ( $P_p$ )  $> 95\%$ ). All the analyses were performed by using CODEML from the PAML v4.9 package [32].

Evolution rate, effective population size dynamics, divergence time and selection pressure were estimated based on two different types of datasets: (i) all sequences of DENV 4-I and (ii) those from different Clades. The spatial diffusion was estimated based on all sequences of DENV 4-I. In order to minimize oversampling of Thailand and Vietnam during the spatial diffusion analysis, we down-sampled dataset for sensitive analysis. The down-sampled dataset included 50 sequences at random, from each Thailand and Vietnam and all the available sequenced strains from other countries/regions, therefore making the sample size 207 sequences.

## Results

### Epidemiological characteristics of DENV 4-I in the world

Fig 1 shows that DENV 4-I were detected in Mainland Southeast Asia and the adjacent provinces of China, Maritime Southeast Asia, the Indian subcontinent, Brazil and Australia.

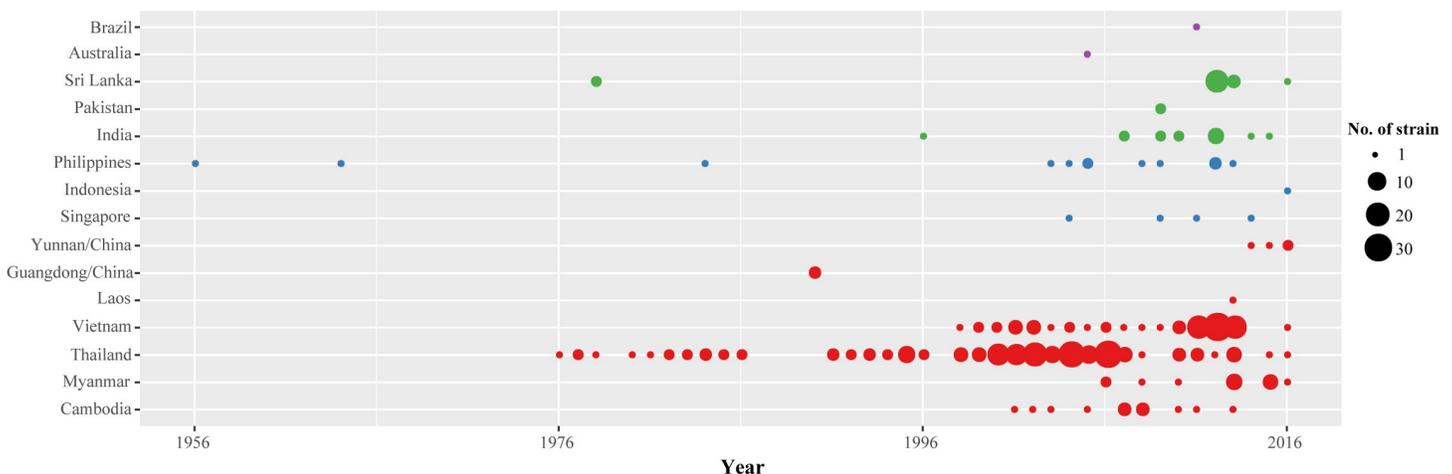
Specifically, the recorded samples revealed that DENV 4-I was mainly observed in Mainland Southeast Asia, especially Thailand and Vietnam. Collection of DENV 4-I covered a period of 60 years. The first strain of DENV 4-I was detected in 1956 in the Philippines, where it transmitted exclusively for some 20 years, according to known reporting and sequencing records. Over the two decades following 1976, most detected strains of DENV 4-I were found to be circulating in Thailand, while a few strains were discovered in other four countries. Detected in a total of 14 countries, DENV 4-I continued to diffuse to more areas around the globe between 1996 and 2016.

Fig 2 shows the evolution and spread of DENV 4-I over time. During the last 60 years, great geographical and genetic diversity has occurred. This is especially prominent during the last two decades after DENV 4-I became more prevalent as shown in the genetic record.

### Phylogenetic and phylogeographical analyses of DENV 4-I

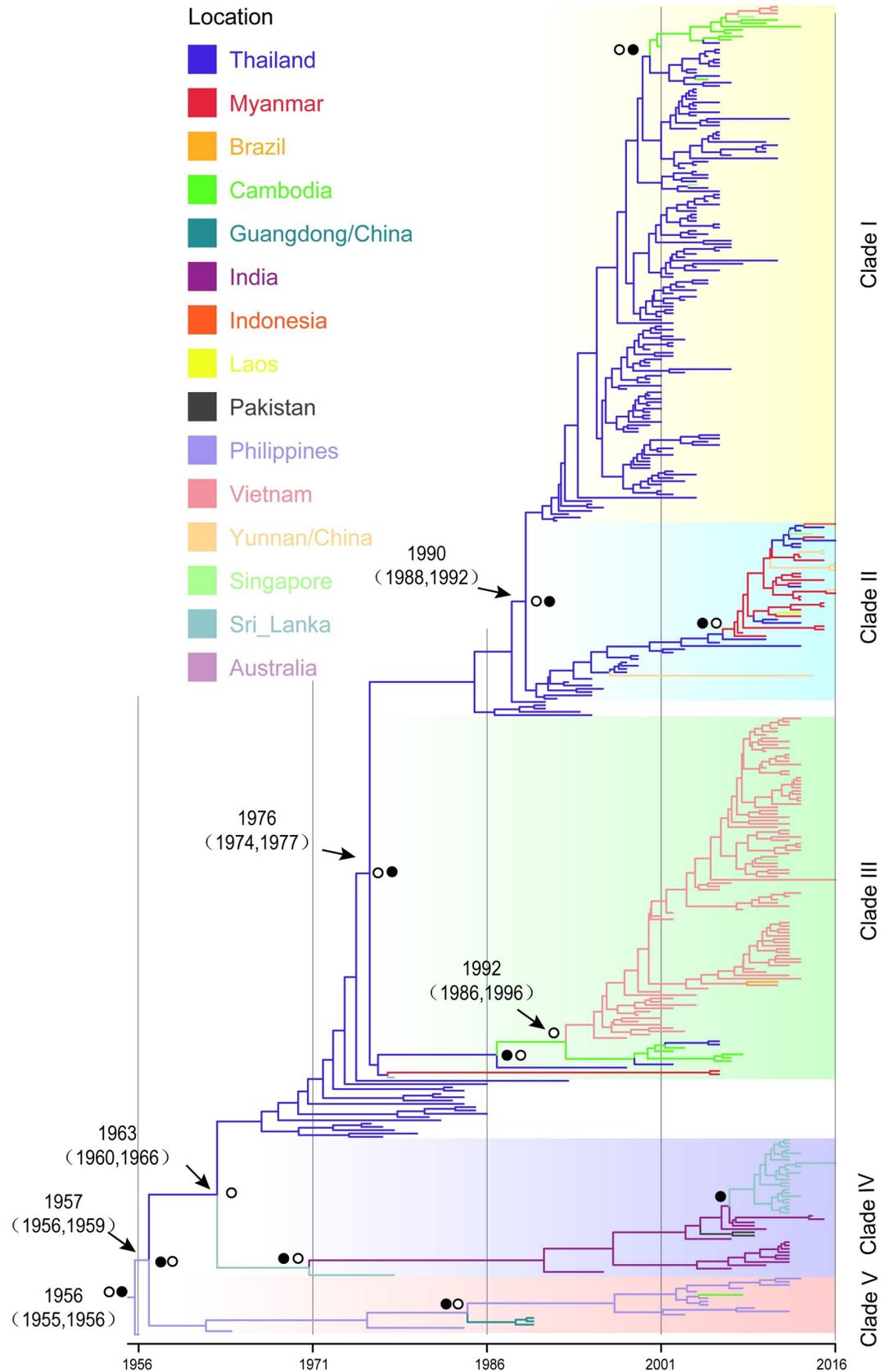
Fig 3 shows the Maximum Clade Credibility (MCC) tree. It indicates that the recent spread of DENV 4-I most likely originated from the Philippines with 0.98 posterior location probability. Viruses evolved in the Philippines and then spread across the sea to Thailand, Cambodia, Australia and China at different times. Thailand played the dominate role in spreading the viruses, gradually spreading virus to the Indian subcontinent, Myanmar, Cambodia, Singapore, Indonesia and China. These viruses diverged around 1957, 1963, 1976 and 1990, and shaped different Clades (Clade I to V). Since the introduction from Thailand, DENV 4-I has evolved in the Indian subcontinent (Clade IV), Myanmar (Clade II) and Cambodia (Clade I and III), respectively. Strains obtained in Vietnam correspond to Clade III, which evolved notably in situ for three decades after introduction from Cambodia. In the Indian subcontinent (Clade IV), dengue virus first apparently arrived in Sri Lanka in the 1960s from Thailand and then spread onto India in the early 1970s. India then became an epicenter for transmission and spread virus to Pakistan and back to Sri Lanka in the 2000s.

Fig 4 shows the detailed spatial diffusion of DENV 4-I as summarized from the MCC tree. The result of the down-sampled dataset showed that the phylogenetic topology and spatial spreading patterns were equivalent with those from full dataset (S2 Fig).



**Fig 2. The temporal distribution of DENV 4-I by location.** The red color dots represents the strains detected in mainland Southeast Asia and its adjacent provinces of China; similarly, blue depicts those found in maritime Southeast Asia; green those from the Indian subcontinent; and purple those few strains found in circulation elsewhere (Brazil and Australia). The diameter size of the circles corresponds to the number unique strains of DENV 4-I in a given year in a given location.

<https://doi.org/10.1371/journal.pntd.0007592.g002>

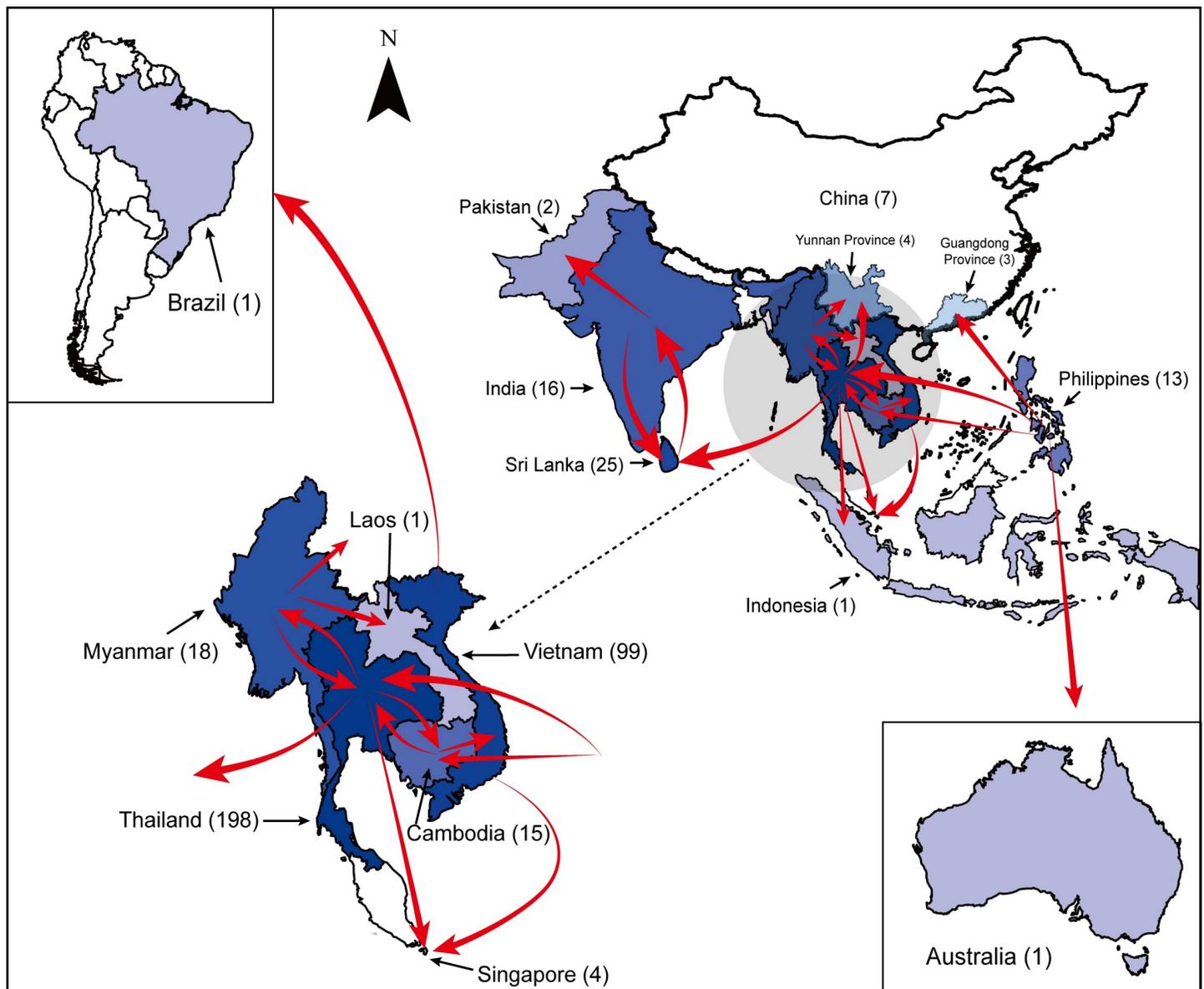


**Fig 3. The Maximum Clade Credibility (MCC) tree summarized for DENV 4-I.** The colors of the branches correspond to their probable geographic location (see the legend). Circles and black dots indicated posterior probability support and ancestral location state probability of main node  $\geq 0.85$ . For the key nodes, the median estimated divergence time and respective 95% HPD intervals were shown.

<https://doi.org/10.1371/journal.pntd.0007592.g003>

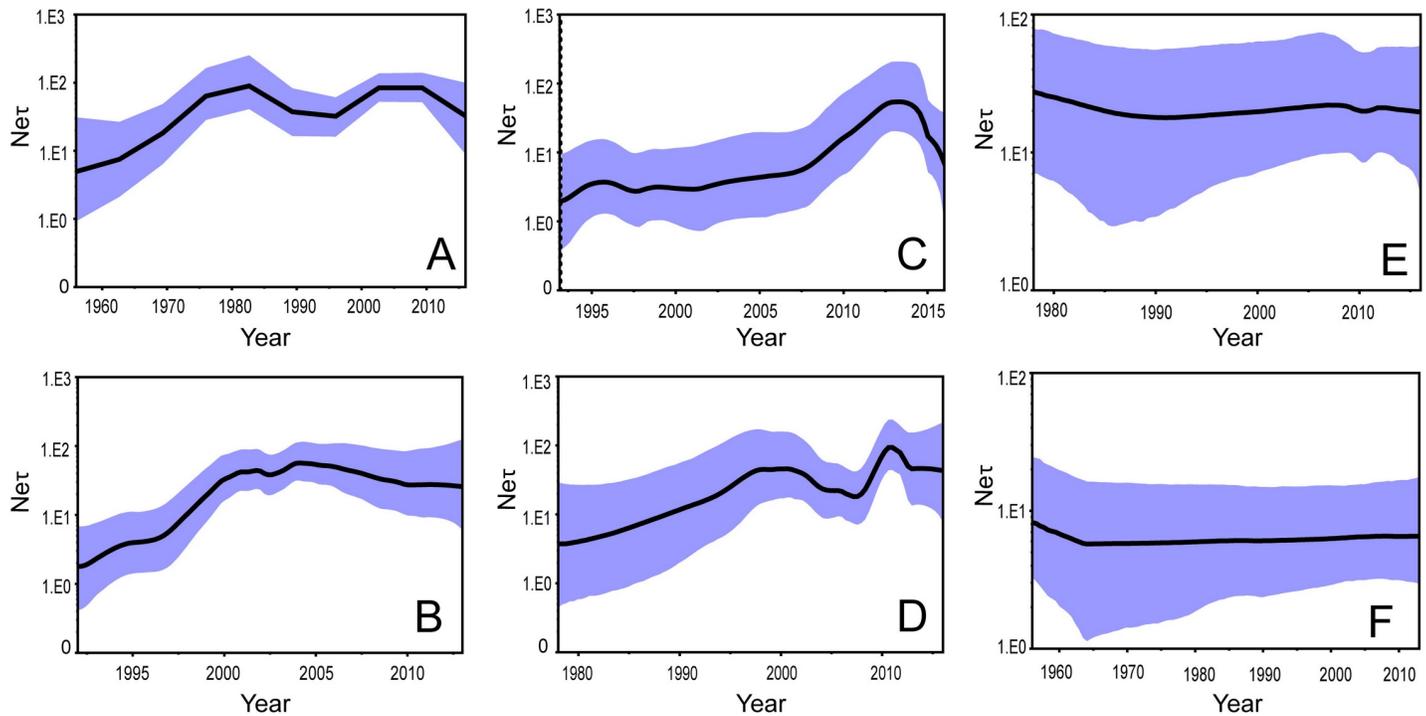
### Demographic history of DENV 4-I and Clades evolution

Fig 5A illustrates the demographic history of DENV 4-I. A fluctuation was observed over the 60 years with an approximately “M” shape. The two highest plateaus were around 1982 and 2006 with a width about 6-years, while the lowest point was around 1996. Analysis of the Clade I dataset revealed, as Fig 5B exhibits, that the effective population size increased approximately



**Fig 4. The global dissemination pathways of DENV 4-I.** Driven by the phylogeographical analysis of all included strains, red arrows indicate the probable direction of trans-border expansion of DENV4-I. The number of unique strains (<100% similarity in the same year) in each location is shown in parenthesis after each location name.

<https://doi.org/10.1371/journal.pntd.0007592.g004>



**Fig 5.** The demographical history of total DENV 4-I (A) and Clade I to V (B-F). The black line represents the median posterior value, and the blue area represents the 95% HPD intervals. The x-axis corresponds to time (years), while the y-axis is the product of the effective population size and the generation in years ( $N_e\tau$ ).

<https://doi.org/10.1371/journal.pntd.0007592.g005>

linearly from 1992 to 2004, and then decreased slowly from 2004 to 2013. The effective population size of Clade II, as Fig 5C conveys, increased slowly for the first two decades and then much more sharply (2007–2013), before a rapid recent decrease (2013–2016). Data from the Clade III in Fig 5D shows that there were two peaks (~2000 and ~2011) over nearly four decades of demographic analysis. The effective population size of Clade IV over the same period decreased slowly and then stayed constant with the inflexion lying around 1990, as Fig 5E depicts. In Fig 5F, one can see the long demographic history pattern of Clade V (1956–2013) is mirrored by Clade IV, but with the inflexion point much earlier (~1965).

### Evolution rate

Table 1 shows the evolution rate and the divergence time of the different Clades’ datasets. The overall evolution rate was  $9.74 \times 10^{-4}$  (95% HPD:  $8.68 \times 10^{-4}$ – $10.82 \times 10^{-4}$ ) nucleotide substitutions/site/year and TMRCA of DENV 4-I was 1956 (95% HPD: 1955–1956) (year). The

**Table 1. Nucleotide substitution rates and divergence times for the E-gene from different datasets.**

Dataset	Nucleotide substitution rate ( $\times 10^{-4}$ substitution/site/year) (95% HPD)	TMRCA (95% HPD, year)
DENV 4-I	9.74 (8.68, 10.82)	1956 (1955, 1956)
Clade I	9.42 (7.78, 11.04)	1991 (1989, 1992)
Clade II	11.3 (8.34, 14.44)	1990 (1988, 1992)
Clade III	8.95 (6.97, 11.06)	1977 (1975, 1978)
Clade IV	9.19 (4.66, 13.46)	1971 (1964, 1976)
Clade V	8.66 (5.90, 11.57)	1962 (1958, 1964)

<https://doi.org/10.1371/journal.pntd.0007592.t001>

mean evolution rate of different Clades was comparable, with the smallest being  $8.66 \times 10^{-4}$  nucleotide substitutions/site/year in Clade V and the largest being  $11.3 \times 10^{-4}$  nucleotide substitutions/site/year in Clade II. TMRCA of Clade I to V was 1991, 1990, 1977, 1971 and 1962, respectively.

### Selective pressure

In the Supplementary Material, [S3 Table](#) shows the summary of positive selection analysis performed on different datasets. M1a vs M2a test and M7 vs M8 test indicated consistently that there was no positive selection across overall DENV 4-I, Clade I, II, IV and V datasets. In Clade III dataset, no positive selection was indicated by M1a vs M2a test, however, M7 vs M8 test inferred a weak positive selection (5.6% of codons with  $\omega = 1.193$ ) with amino acid site 95  $P_p > 0.95$ .

### Discussion

In this study, using the most comprehensive and largest dataset(s) available we mapped the spatial distribution and 60-years evolution of one genotype from one of the four serotypes of dengue virus, DENV 4-I. We further separated this large dataset into different Clades and analyzed the evolutionary dynamics of Clades separately. Our study shows that the spatial distribution of DENV 4-I is mainly restricted to Southeast Asia and the Indian subcontinent. The recent spread of DENV 4-I likely originated from Southeast Asia—namely the Philippines, from where it then spread all the way to the Indian subcontinent, Australia and Latin America. DENV 4-I evolved in situ in Southeast Asia and on the Indian subcontinent. Although DENV 4-I cases occasionally were found elsewhere, this genotype didn't undergo in situ evolution and largely failed to establish. We found that mainland Southeast Asia, specifically Thailand, was at the center of the global spread of the viruses, which in time contributed to the observed diversity.

Although Thailand's strains dominated the records, the Philippines' strains were at the basal location of the phylogeny. The analyzed diverge time (95% HPD: 1955–1956) is consistent with the timing of DENV 4-I virus being first detected in the Philippines. Therefore, we propose that global spread of DENV 4-I originated very likely from the Philippines. From the Philippines, DENV 4-I spread to Thailand and then on to nearby countries including Sri Lanka. On this point our findings clash with one previous study's results indicating that DENV 4-I originated from Thailand, from where it spread to the neighboring countries of the Philippines and Sri Lanka [17].

We found that the strains of DENV 4-I—monophyletic Clade IV, detected in the Indian subcontinent including Sri Lanka, India and Pakistan, probably originated in Thailand, then evolved in situ, but as of 2016, had not yet spread outside of the Indian subcontinent. Dengue is endemic in the Indian subcontinent and four serotypes have been co-circulating, with DENV 2 and DENV 3 dominating [33]. However, the distribution and dissemination patterns of the different genotypes are not uniform. A study showed that the DENV 3 genotype III (DENV 3-III) was spread from the Indian subcontinent to East Africa first and then from there to Latin America [34]. Another study on DENV 2 showed that the cosmopolitan genotype was spread from the Indian subcontinent directly to Latin America [14]. In addition to the Indian subcontinent, dengue is also endemic in Brazil. Even in two separate regions of the world where dengue is endemic, the specific evolutionary drivers of a given serotype and genotype may differ.

DENV4-I was introduced to Brazil in 2011. Our study indicated that the strain was probably imported from mainland Southeast Asia, which was consistent with the analysis conducted

in a previous study [35]. Curiously, DENV 4-I apparently disappeared quickly from Brazil after the introduction. This is different from the expansion/establishment patterns of DENV 4-II, which has become established in the Caribbean since early 1980s after being introduced from Southeast Asia [36]. Human movement is known to play a significant role in contributing to the virus distribution at large spatial scales (e.g., national, international) [37], because of the limited range of mosquitos' flying distance. However, human movement alone cannot explain the different expansion pathways and establishment of new genotypes, for example, in the DENV 4-I dynamics on the Indian subcontinent and Brazil even after controlling for the climate suitability and mosquito abundance. These studies show that the spread of dengue virus is not merely shaped by human movement. Human movement is a necessary but not sufficient condition. Different pathways of genotypes spread may be associated with the strain virulence, human population immunity, previous exposure to other dengue viruses, in addition to the volume and timing of human population movements, the local environment, and numerous other factors.

Our study on evolution rate showed that the substitution rate of DENV 4-I is 9.74 (95% HPD: 8.68–10.82)  $\times 10^{-4}$  nucleotide substitutions/site/year. This is comparable with the value found by Klungthong *et al.* [38] (10.72 (95% HPD: 8.41–13.11)  $\times 10^{-4}$  nucleotide substitutions/site/year), but is approximately twice that of the one estimated by Twiddy *et al.* (5.42  $\times 10^{-4}$  substitutions/site/year) using an admittedly much smaller dataset [39]. The range of mean evolution rate from DENV 2–4 is found to be (8.3–10.7)  $\times 10^{-4}$  nucleotide substitutions/site/year [16], which included Asian I and Asian-American genotype of DENV 2, genotype I–III of DENV 3, DENV 4-II. Although DENV 4-I was shown to evolve with a relatively large substitution rate comparing to some other genotypes [16], there is no significant rate differences among the different dengue serotypes. Among the five Clades of DENV 4-I, we found that the highest evolution rate was among Clade II strains (Table 1). The MCC tree (Fig 3) indicates that the strains of Clade II spread rapidly in mainland Southeast Asia. The rapid spread and the subsequent replication in a large susceptible population of hosts could account for the higher rate of evolution, when compared to the other Clades in DENV 4-I having relatively smaller pools of susceptible hosts. We found no evidence of positive selection in most of the Clades of DENV 4-I, although there was evidence for weak positive selection in Clade III. Some studies have suggested purifying selection [38], where majority of amino acid changes within infected hosts, are deleterious in the long run and are eventually removed from the population [40]. This purifying selection induces the distinct ladder-like phylogeny. Our study confirms this purifying selection process as the predominant evolutionary force acting on DENV 4-I.

Our study also showed that DENV 4-I experienced a fluctuating demographic pattern although occurring in a low prevalence compared to other serotypes. The fluctuation of effective infected population size might be shaped by human population immunity/susceptibility. DENV 4-I was first detected in the Philippines and spread to Thailand and Sri Lanka before 1980, where humans were lacking previous exposure and thus immunity. The effective infected population size decreased once immunity was built up as result of previous exposure. DENV 4-I viruses spread onto Vietnam around 1992 (95% HPD: 1986–1996) and evolved in situ, which might have induced the secondary increase. This observation agrees with the study by Villabona-Arenas *et al.* using the dataset from 1956 to 2008. They indicated that the effective population size was estimated to be the largest at two time periods around 1982 and 2005 [17], which is similar to our result shown in Fig 5A using a larger dataset covering a longer period.

Dengue is endemic in Vietnam. Most of the DENV 4-I strains detected in Vietnam shaped the Clade III, as the in situ evolution occurred since DENV 4-I started circulating in Vietnam

in 1998. It did not dominate during dengue epidemics until 2013 when a large outbreak was occurred having 204,661 clinical cases with nearly 50% of them having DENV 4-I in central Vietnam [10]. This occurrence coincided with the demographic history of Clade III. Lacking immunity to DENV 4-I among the Vietnamese human population could be a reason to cause second infection and therefore severe dengue which resulted in being more likely reported. Amino acid 95 under positive selection could also account for this large outbreak. The E protein ectodomain can be divided into three structural domains designated domain I-III. Domain II contains the fusion loop (residue 98–111), which interacts with the host endosomal membrane, leading to virus-mediated membrane fusion and allowing the newly infecting virus to initiate the cellular replication cycle. As residue 95 is located three residues downstream of the fusion loop, it is likely to indirectly affect the process of membrane fusion. The Vietnam strains experiencing positive selection pressure did not spread outside of Vietnam based on the available data and our findings. Enhanced surveillance to these strains could be very helpful to aid in understanding and controlling this potentially devastating virus strain.

Although our study represents the most comprehensive study on the evolutionary dynamics of DENV 4-I using the largest dataset available to date, the results should be interpreted cautiously given various limitations. For example, itself, reporting of DENV, especially in the genetic record, is a source of many types of potential bias, of particular concern in locations with limited resources for virologic diagnostic and reporting capacity. While our dataset represents an opportunistic, but highly useful sample of occurrence, it should be noted that we worked with a limited subset of data, rather than a complete record of global DENV 4-I transmission dynamics. To account for sampling bias, in this study we constructed and analyzed a down-sampled dataset for sensitive analysis, finding an equivalent result to that which was obtained with the full dataset. Nevertheless, the quality of this type of study will be increased greatly with an enhanced global dengue surveillance, greater access to next generation diagnostic and sequencing tools, and improved data sharing systems. Our current study can be described as exploratory research, as this is the first time that the geographic spread and evolutionary dynamics of the DENV 4-I was mapped out and analyzed based on a large dataset. Given that other dengue research in this area is sparse and rarely conducts genotype-specific analysis, our targeted focus on just one genotype of DENV 4 makes comparison difficult. We hope that our study can stimulate further research studies in this area so that in the future researchers can compare different genotypes and understand better the similarities and differences among them. Our mapping of the genotypes' distribution pattern may help to generate hypotheses on the specific mechanisms mediating the spread of DENV 4-I. This understanding is potentially of great utility in the generation of health policies and practices on dengue prevention and control.

## Conclusion

In this study, we have investigated the global patterns of DENV 4-I dissemination—its spatial and temporal distribution. This is the most extensive molecular epidemiological study of DENV 4 genotype I to date to our knowledge. Our results indicate that recent spread of DENV 4-I originated in maritime Southeast Asia, probably from the Philippines, from where it spread to mainland Southeast Asia, and then on to the Indian subcontinent. Thailand acted as a distribution hub for spreading the virus regionally and globally. Within the India subcontinent, India was the distribution center for spreading the virus. We found that there is no uniform spreading pattern among genotypes. In addition, purifying selection was still the dominant acting force on E gene to shape the evolution, but weak positive selection existed in dengue viruses detected in Vietnam.

This work is a first step towards increased understanding of the underlying mechanisms governing the spread of DENV 4-I virus. Our study suggests that surveillance could be enhanced to better leverage next generation sequencing for informing dengue control practices. Regional cooperation should be strengthened to determine and communicate information on the genotype-specific spreading pathways, to explore the related underlying mechanisms, and ultimately to better coordinate dengue control efforts globally.

## Supporting information

**S1 Fig.** Results of substitution saturation analysis (A) and likelihood mapping analysis (B). (TIF)

**S2 Fig. Maximum Clade Credibility (MCC) tree summarized for down-sampled dataset of DENV 4-I.** The colors of the branches corresponded to their probable geographic location (see the legend). The number was the ancestral location probability of key node. (TIF)

**S1 Table. DENV4 genotype I.** Strains from GenBank were included in the analyses. (XLSX)

**S2 Table. Model selection based on marginal likelihood estimates.** (XLSX)

**S3 Table. Parameters estimated from selection pressure analyses.** Here positive selection was detected based on different datasets. (XLSX)

## Author Contributions

**Conceptualization:** Shaowei Sang, Qiyong Liu.

**Data curation:** Shaowei Sang.

**Formal analysis:** Shaowei Sang.

**Funding acquisition:** Shaowei Sang.

**Investigation:** Hongning Zhou, Xiaofang Guo.

**Supervision:** Shaowei Sang, Qiyong Liu.

**Writing – original draft:** Shaowei Sang.

**Writing – review & editing:** Jing Liu-Helmersson, Mikkel B. M. Quam, Haixia Wu, Qiyong Liu.

## References

1. Kyle JL, Harris E. Global spread and persistence of dengue. *Annual review of microbiology*. 2008; 62:71–92. <https://doi.org/10.1146/annurev.micro.62.081307.163005> PMID: 18429680.
2. Guzman MG, Harris E. Dengue. *Lancet*. 2015; 385(9966):453–65. [https://doi.org/10.1016/S0140-6736\(14\)60572-9](https://doi.org/10.1016/S0140-6736(14)60572-9) PMID: 25230594.
3. Vos T, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*. 2017; 390(10100):1211–59.

4. Shepard DS, Undurraga EA, Halasa YA, Stanaway JD. The global economic burden of dengue: a systematic analysis. *The Lancet Infectious diseases*. 2016; 16(8):935–41. [https://doi.org/10.1016/S1473-3099\(16\)00146-8](https://doi.org/10.1016/S1473-3099(16)00146-8) PMID: 27091092.
5. Chen R, Vasilakis N. Dengue—quo tu et quo vadis? *Viruses*. 2011; 3(9):1562–608. <https://doi.org/10.3390/v3091562> PMID: 21994796; PubMed Central PMCID: PMC3187692.
6. Hammon WM, Rudnick A, Sather GE. Viruses associated with epidemic hemorrhagic fevers of the Philippines and Thailand. *Science*. 1960; 131(3407):1102–3. <https://doi.org/10.1126/science.131.3407.1102> PMID: 14399343.
7. Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, Duda KA, et al. Global spread of dengue virus types: mapping the 70 year history. *Trends in microbiology*. 2014; 22(3):138–46. <https://doi.org/10.1016/j.tim.2013.12.011> PMID: 24468533; PubMed Central PMCID: PMC3946041.
8. Gaunt MW, Sall AA, de Lamballerie X, Falconar AK, Dzihivianian TI, Gould EA. Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. *The Journal of general virology*. 2001; 82(Pt 8):1867–76. <https://doi.org/10.1099/0022-1317-82-8-1867> PMID: 11457992.
9. Nisalak A, Endy TP, Nimmannitya S, Kalayanarooj S, Thisayakorn U, Scott RM, et al. Serotype-specific dengue virus circulation and dengue disease in Bangkok, Thailand from 1973 to 1999. *The American journal of tropical medicine and hygiene*. 2003; 68(2):191–202. PMID: 12641411.
10. Takamatsu Y, Nabeshima T, Nguyen TT, Dang DT, Pham LH, Pham HT, et al. A Dengue virus serotype 4-dominated outbreak in central Vietnam, 2013. *Journal of clinical virology: the official publication of the Pan American Society for Clinical Virology*. 2015; 66:24–6. <https://doi.org/10.1016/j.jcv.2015.02.016> PMID: 25866331.
11. Ngwe Tun MM, Kyaw AK, Makki N, Muthugala R, Nabeshima T, Inoue S, et al. Characterization of the 2013 dengue epidemic in Myanmar with dengue virus 1 as the dominant serotype. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2016; 43:31–7. <https://doi.org/10.1016/j.meegid.2016.04.025> PMID: 27154331.
12. Uehara A, Tissera HA, Bodinayake CK, Amarasinghe A, Nagahawatte A, Tillekeratne LG, et al. Analysis of Dengue Serotype 4 in Sri Lanka during the 2012–2013 Dengue Epidemic. *The American journal of tropical medicine and hygiene*. 2017; 97(1):130–6. <https://doi.org/10.4269/ajtmh.16-0540> PMID: 28719296; PubMed Central PMCID: PMC5508889.
13. Sang S, Chen B, Wu H, Yang Z, Di B, Wang L, et al. Dengue is still an imported disease in China: a case study in Guangzhou. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2015; 32:178–90. <https://doi.org/10.1016/j.meegid.2015.03.005> PMID: 25772205.
14. Wei K, Li Y. Global evolutionary history and spatio-temporal dynamics of dengue virus type 2. *Scientific reports*. 2017; 7:45505. <https://doi.org/10.1038/srep45505> PMID: 28378782; PubMed Central PMCID: PMC5381229.
15. Villabona-Arenas CJ, Zannotto PM. Worldwide spread of Dengue virus type 1. *PloS one*. 2013; 8(5): e62649. <https://doi.org/10.1371/journal.pone.0062649> PMID: 23675416; PubMed Central PMCID: PMC3652851.
16. Araujo JM, Nogueira RM, Schatzmayr HG, Zannotto PM, Bello G. Phylogeography and evolutionary history of dengue virus type 3. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2009; 9(4):716–25. <https://doi.org/10.1016/j.meegid.2008.10.005> PMID: 19010450.
17. Villabona-Arenas CJ, Zannotto PM. Evolutionary history of Dengue virus type 4: insights into genotype phylogenetics. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2011; 11(5):878–85. <https://doi.org/10.1016/j.meegid.2011.02.007> PMID: 21335103.
18. Guo X, Yang H, Wu C, Jiang J, Fan J, Li H, et al. Molecular Characterization and Viral Origin of the First Dengue Outbreak in Xishuangbanna, Yunnan Province, China, 2013. *The American journal of tropical medicine and hygiene*. 2015; 93(2):390–3. <https://doi.org/10.4269/ajtmh.14-0044> PMID: 26078324; PubMed Central PMCID: PMC4530767.
19. Katoh K, Asimenos G, Toh H. Multiple alignment of DNA sequences with MAFFT. *Methods in molecular biology*. 2009; 537:39–64. [https://doi.org/10.1007/978-1-59745-251-9\\_3](https://doi.org/10.1007/978-1-59745-251-9_3) PMID: 19378139.
20. Martin DP. Recombination detection and analysis using RDP3. *Methods in molecular biology*. 2009; 537:185–205. [https://doi.org/10.1007/978-1-59745-251-9\\_9](https://doi.org/10.1007/978-1-59745-251-9_9) PMID: 19378145.
21. Xia X. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Molecular biology and evolution*. 2013; 30(7):1720–8. <https://doi.org/10.1093/molbev/mst064> PMID: 23564938; PubMed Central PMCID: PMC3684854.

22. Strimmer K, von Haeseler A. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences of the United States of America*. 1997; 94(13):6815–9. <https://doi.org/10.1073/pnas.94.13.6815> PMID: 9192648; PubMed Central PMCID: PMC21241.
23. Schmidt HA, Strimmer K, Vingron M, von Haeseler A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*. 2002; 18(3):502–4. <https://doi.org/10.1093/bioinformatics/18.3.502> PMID: 11934758.
24. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*. 2007; 7:214. <https://doi.org/10.1186/1471-2148-7-214> PMID: 17996036; PubMed Central PMCID: PMC2247476.
25. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*. 2012; 9(8):772. <https://doi.org/10.1038/nmeth.2109> PMID: 22847109; PubMed Central PMCID: PMC4594756.
26. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular biology and evolution*. 2008; 25(7):1459–71. <https://doi.org/10.1093/molbev/msn090> PMID: 18408232; PubMed Central PMCID: PMC3302198.
27. Raftery A, Newton M, Satagopan J, Krivitsky P. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. In: Bernardo J, Bayarri M, Berger J, editors. *Bayesian Statistics*: Oxford: Oxford University Press; 2007. p. 1–45.
28. Suchard MA, Weiss RE, Sinsheimer JS. Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular biology and evolution*. 2001; 18(6):1001–13. <https://doi.org/10.1093/oxfordjournals.molbev.a003872> PMID: 11371589.
29. Newton MA, Raftery AE. Approximate Bayesian Inference with the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society*. 1994; 56(1):3–48.
30. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular biology and evolution*. 2012; 29(9):2157–67. <https://doi.org/10.1093/molbev/mss084> PMID: 22403239; PubMed Central PMCID: PMC3424409.
31. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. *Trends in ecology & evolution*. 2000; 15(12):496–503. PMID: 11114436.
32. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer applications in the biosciences: CABIOS*. 1997; 13(5):555–6. PMID: 9367129.
33. Raheel U, Faheem M, Riaz MN, Kanwal N, Javed F, Zaidi N, et al. Dengue fever in the Indian Subcontinent: an overview. *Journal of infection in developing countries*. 2011; 5(4):239–47. PMID: 21537064.
34. Messer WB, Gubler DJ, Harris E, Sivananthan K, de Silva AM. Emergence and global spread of a dengue serotype 3, subtype III virus. *Emerging infectious diseases*. 2003; 9(7):800–9. <https://doi.org/10.3201/eid0907.030038> PMID: 12899133; PubMed Central PMCID: PMC3023445.
35. Nunes MR, Faria NR, Vasconcelos HB, Medeiros DB, Silva de Lima CP, Carvalho VL, et al. Phylogeography of dengue virus serotype 4, Brazil, 2010–2011. *Emerging infectious diseases*. 2012; 18(11):1858–64. <https://doi.org/10.3201/eid1811.120217> PMID: 23092706; PubMed Central PMCID: PMC3559147.
36. Bennett SN, Holmes EC, Chirivella M, Rodriguez DM, Beltran M, Vorndam V, et al. Selection-driven evolution of emergent dengue virus. *Molecular biology and evolution*. 2003; 20(10):1650–8. <https://doi.org/10.1093/molbev/msg182> PMID: 12832629.
37. Stoddard ST, Morrison AC, Vazquez-Prokopec GM, Paz Soldan V, Kochel TJ, Kitron U, et al. The role of human movement in the transmission of vector-borne pathogens. *PLoS neglected tropical diseases*. 2009; 3(7):e481. <https://doi.org/10.1371/journal.pntd.0000481> PMID: 19621090; PubMed Central PMCID: PMC2710008.
38. Klungthong C, Zhang C, Mammen MP Jr., Ubol S, Holmes EC. The molecular epidemiology of dengue virus serotype 4 in Bangkok, Thailand. *Virology*. 2004; 329(1):168–79. <https://doi.org/10.1016/j.virol.2004.08.003> PMID: 15476884.
39. Twiddy SS, Holmes EC, Rambaut A. Inferring the rate and time-scale of dengue virus evolution. *Molecular biology and evolution*. 2003; 20(1):122–9. <https://doi.org/10.1093/molbev/msg010> PMID: 12519914.
40. Holmes EC. Patterns of intra- and interhost nonsynonymous variation reveal strong purifying selection in dengue virus. *Journal of virology*. 2003; 77(20):11296–8. <https://doi.org/10.1128/JVI.77.20.11296-11298.2003> PMID: 14512579; PubMed Central PMCID: PMC224983.