

# Evolution, heterogeneity and global dispersal of cosmopolitan genotype: The most widespread genotype of Dengue Virus type 2

**Surya Pavan Yenamandra**

National Environment Agency

**Carmen Koo**

National Environment Agency

**Suzanna Chiang**

National Environment Agency

**Lim Han Shi Jeri**

Nanyang Technological University

**Zhen Yuan Yeo**

National University of Singapore

**Lee Ching Ng**

National Environment Agency

**Hapuarachchige Chanditha Hapuarachchi** (✉ [Chanditha\\_hapuarachchi@nea.gov.sg](mailto:Chanditha_hapuarachchi@nea.gov.sg))

National Environment Agency

---

## Research Article

**Keywords:** Dengue virus, genotype, evolution, phylogeography, dispersal

**Posted Date:** April 7th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-389331/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Dengue virus type 2 (DENV-2) contributes substantially to the dengue burden and dengue-related mortality in the tropics and sub-tropics. DENV-2 includes six genotypes, among which cosmopolitan genotype is the most widespread. The present study investigated the evolution, intra-genotype heterogeneity and dispersal of cosmopolitan genotype to understand unique genetic characteristics that have shaped the molecular epidemiology and distribution of cosmopolitan lineages. The spatial analysis demonstrated a wide geo-distribution of cosmopolitan genotype through an extensive inter-continental network, anchored in Southeast Asia and Indian sub-continent. Intra-genotype analyses using 2,392 envelope gene sequences revealed six distinct lineages within the cosmopolitan genotype, namely the Indian sub-continent lineage and five other lineages. Indian sub-continent lineage was the most diverged among six lineages and has almost reached the nucleotide divergence threshold of 6% within *E* gene to qualify as a separate genotype. Genome wide amino acid signatures and selection pressure analyses further suggested differences in evolutionary characteristics between the Indian sub-continent lineage and other lineages. The present study narrates a comprehensive genomic analysis of cosmopolitan genotype and presents notable genetic characteristics that occurred during its evolution and global expansion. Whether those characteristics conferred a fitness advantage to cosmopolitan genotype in different geographies warrant further investigations.

## Highlights

- Cosmopolitan genotype demonstrated inter-continental dispersal through an extensive spatial network.
- Genome analyses classified cosmopolitan genotype into six lineages – Indian sub-continent lineage and five other lineages, reflecting high intra-genotype heterogeneity and wide geographical presence.
- Indian sub-continent lineage is the most divergent among six lineages and showed a common ancestry with genome sequences reported from Indonesia in mid-1970s.
- *E* gene nucleotide divergence data suggested that Indian sub-continent lineage is maturing into a new genotype.
- Genome wide amino acid signatures and selection pressure analyses testified differences in evolutionary characteristics between the Indian sub-continent lineage and other lineages.

## Introduction

Dengue is the most common arthropod-borne acute febrile illness, caused by Dengue virus (DENV). Dengue is endemic in the tropics and subtropics, where approximately half of the human population lives at risk [1]. The annual dengue burden worldwide has increased drastically during the last four decades, especially after 1990s [2, 3]. DENV is expanding its global footprint amidst environmental and climatic changes that promote the propagation and distribution of its vectors, *Aedes aegypti* and *Aedes albopictus*.

DENV is categorized into four serotypes (DENV-1-4) based on genetic and antigenic characteristics. The inter-serotype nucleotide variability is ~ 30% [4, 5]. Each serotype is genetically sub-divided into genotypes [6, 7]. Different genotypes show variable adaptability that accompanies genetic modifications in DENV, which may enhance virus dispersal and epidemic potential [8, 9] in different geographies. Among four serotypes, DENV-1 and DENV-2 are the most common types associated with known outbreaks [3]. According to empirical data, the highest pooled mortality rate has been reported during DENV-2 outbreaks [10]. Studies have also shown that DENV-2 causes more severe secondary infections than other serotypes [11, 12]. These evidences indicate an important role played by DENV-2 in dengue epidemiology and global health burden.

DENV-2 includes five non-sylvatic genotypes (Asian I, Asian II, American, Asian-American and Cosmopolitan) that are dominant in different regions [13]. Asian I and Asian II genotypes are present in the Asian continent. The already-extinct American genotype was largely restricted to Central and South America. Asian-American genotype circulates in Central and South America as well as in Southeast Asia [14]. Cosmopolitan genotype circulates in the Asia-Pacific, Middle East and Africa and is the most widespread genotype of DENV-2 [15–22]. It has expanded further into Uganda, Bhutan, Borneo, Brunei, Japan, Maldives, Micronesia and Tanzania in recent years [23–33]. Given its wide geographical presence, cosmopolitan genotype contributes substantially to the global DENV-2 burden.

Based on current understanding, we hypothesize that DENV-2 cosmopolitan genotype has evolved in parallel to its global dispersal to achieve a higher heterogeneity than other DENV-2 genotypes. The current study, therefore, investigated the evolutionary characteristics and global dispersal patterns of DENV-2 cosmopolitan genotype associated with its genetic heterogeneity, distinctness and geographical dominance. The findings demonstrated that cosmopolitan genotype is spread through an extensive spatial network that is characterized by the dominance of distinct intra-genotype lineages in different geographies.

## Materials And Methods

### Sequence data

Two datasets were used in the analyses. The envelope (*E*) gene dataset included 4,551 complete *E* gene sequences (1,485 bp) of DENV-2, comprising of 3,598 sequences retrieved from the NCBI nucleotide repository and 1,048 sequences newly generated at the Environmental Health Institute, Singapore. These new sequences included those obtained from patient sera (n=915) and field-caught mosquitoes (n=133). They were generated as part of the arbovirus surveillance programme in Singapore [34, 35]. Among 4,551 *E* gene sequences, 2,392 belonged to the cosmopolitan genotype. The second dataset included complete genomes (n=1,548) of DENV-2 retrieved from the NCBI nucleotide repository, and 158 newly generated complete genomes of cosmopolitan genotype. Sequences of each dataset were aligned by using the MAFFT software [36].

### Accordance to relevant guidelines

All DENV-positive patient sera used for the generation of new *E* gene sequences (n=915) were collected after obtaining the written informed consent from respective patients. All sera were utilized in accordance with the guidelines approved by the Institutional Review Board of National Environmental Agency, Singapore (IRB003.1).

### **Ethical approval**

The study was approved by the Institutional Review Board of National Environmental Agency, Singapore (IRB003.1).

### **Sequencing of DENV-2 envelope gene and complete genomes**

Envelope gene sequences (n=1,048) were generated directly from patient sera and mosquito homogenates. Complete genomes (n=158) were generated from isolates obtained from patient sera. DENV was isolated from sera using the *Ae. albopictus* C6/36 mosquito cell line (ATCC CRL-1660) as described in detail elsewhere [37]. Viral RNA was extracted from sera using the QIAGEN QIAamp viral RNA mini kit (QIAGEN, Hilden, Germany) according to manufacturer's recommendations. Viral RNA from mosquitoes was extracted as described previously [38]. DENV-2 typing, synthesis of complementary DNA (cDNA) and *E* gene and complete genome sequencing were done as described elsewhere [34, 37]. Amplification and sequencing primers used are given in Supplementary Table 1. Raw nucleotide sequences were assembled using the Lasergene package version 15.0 (DNASTAR Inc., Madison, WI, USA).

### **Spatial distribution of DENV-2 genotypes**

Information pertaining to the country of origin (location) and year of isolation (time) of each sequence was extracted from NCBI flat files using a python script. Sylvatic genotype was excluded from the analyses because of extremely low numbers reported and its negligible contribution to global dengue burden. All *E* gene sequences (n=4,551) were categorized into known genotypes based on a neighbor-joining phylogenetic tree constructed by using the maximum-likelihood composite model in MEGA7 software suite [39]. A database (.csv) was prepared for each genotype to include the corresponding country of origin as well as latitude and longitude information of the capital city of each country. To determine the extent of global presence, countries that have reported each genotype, except the American genotype, were ranked into >90th, 75-90th, 60-74th, 30-59th and <30th percentile categories based on the number of sequences reported. Due to low sample size, countries that have reported the American genotype were categorized into >90<sup>th</sup>, >89-80<sup>th</sup>, >79-70<sup>th</sup>, >69-60<sup>th</sup> and <60<sup>th</sup> percentiles for better visualization. The data for each genotype was plotted onto the global map by using a python script containing cartopy.mpl.gridliner map builder module.

### **Bayesian phylogeny of DENV-2 complete genomes**

Bayesian phylogeny analysis was conducted to determine the evolutionary characteristics of DENV-2 genotypes using 419 complete genome sequences. The analysis included all cosmopolitan genotype sequences (n=380) and subsets of Asian I (n=10), Asian II (n=10), Asian-American (n=8), American (n=8) and sylvatic (n=3) genotypes. The sequences of “non-cosmopolitan” genotypes were selected to represent the time and diversity span of each genotype determined based on a neighbor-joining tree constructed in MEGA7 software suite [39]. Each sequence was time tagged based on the year of isolation information. The Bayesian phylogeny was constructed by using general time reversible substitution model with gamma distribution and invariant sites (GTR+ $\Gamma$ 5+ I), codon partition (1+2) and 3, uncorrelated log-normal relaxed molecular clock model and Bayesian skyline prior (stepwise, 10 bins) in BEAST v1.10.4 software suite [40, 41]. The best fitting nucleotide model for the analyzed dataset was determined by using jModeltest 2.1.5 [42]. The BEAST run included 400 million Markov Chain Monte Carlo (MCMC) chains that were sampled at every 40,000 generations to ensure convergence at a minimum effective sample size (ESS) of >200. The MCMC output was analysed using Tracer v.1.6 software (available at <http://beast.bio.ed.ac.uk/Tracer>). The maximum clade credibility (MCC) tree was constructed using TreeAnnotator v.1.10.4 software and visualized in FigTree v.1.4.4 software (available at <http://tree.bio.ed.ac.uk/software/figtree/>). Evolutionary parameters such as evolutionary rates and time to most recent common ancestor (tMRCA) were calculated based on the MCC tree. The uncertainty in parameter estimates was expressed as 95% highest posterior density (HPD).

### **Discrete phylogeography analysis**

Dispersal pattern of DENV-2 cosmopolitan genotype was determined by using the Bayesian Stochastic Search Variable studies (BSSVS) analysis [43]. Each country was treated as a discrete trait in the analysis. To minimize the redundancy and computational resources required, identical sequences reported from each country in a particular year were removed from the original *E* gene alignment. The final alignment included 929 *E* gene sequences, reported from 27 countries. BSSVS analysis was conducted by using GTR+ $\Gamma$ 5+ I as the best-fitting substitution model based on jModelTest output [42], Bayesian skyline demographic prior with default 10 bins and a relaxed uncorrelated lognormal clock [44] in a MCMC run of 100 million generations, sampling every 10,000 states. Phylogeographic reconstructions and Bayes factor (BF) calculations were done in SPREAD 1.0.4 software [45] by comparing the posterior and prior probability of individual rates to test the significant linkage between locations. Links with BF values >3 and posterior probability  $\geq$ 0.75 were considered well-supported. Well-supported links were categorized into very strong (BF>30) and decisive (BF>100) [43, 46].

### **Mutation profiling of cosmopolitan genotype sequences**

Unique mutations of the cosmopolitan genotype were identified by using an alignment of 1,548 complete coding sequences of DENV-2 retrieved from NCBI nucleotide repository. The analyses were done in Jalview 2.11.1.3 software [47]. Each sequence was categorized into a known genotype based on phylogenetic clustering. Firstly, residues that were conserved in >96% of cosmopolitan genotype sequences, but different in other genotypes of DENV-2 were identified as unique substitutions of the

cosmopolitan genotype. Secondly, the analysis zoomed into cosmopolitan genotypes sequences to identify unique substitutions of the Indian sub-continent lineage as compared to other lineages in the same manner. Each mutation signature comprised the combination of all such unique substitutions.

### **Genetic distance analyses of DENV-2 sequences**

The overall nucleotide similarity and pairwise nucleotide diversity ( $\pi$  index) were determined between the cosmopolitan genotype (Indian sub-continent and other lineages) and other genotypes of DENV-2 using the complete *E* gene (1,485 bp) alignment. The overall nucleotide similarity among different categories/genotypes were calculated in BioEdit version 7.2.5 [48]. The  $\pi$  index is defined as the number of nucleotide differences between two sequences per base pair [49]. The  $\pi$  index was determined in Biopython [50] by estimating the pairwise nucleotide diversity of each pair of sequences calculated with Dendropy [51], considering cosmopolitan (other lineage) sequences as the reference dataset.

### **Selection pressure analyses**

The selection pressure acting on each codon of complete coding sequences of 488 (99 Indian sub-continent lineage and 389 other lineages) were measured by the ratio of non-synonymous to synonymous rates computed in HyPhy open-source software package as implemented in Datamonkey web-server (Delpont et al., 2010). Single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), mixed effect model of evolution (MEME) and fast unbiased Bayesian approximation (FUBAR) methods were used to estimate the sites under selection. The GTR nucleotide substitution bias model and neighbor joining phylogeny were used as analytical parameters in different approaches. An integrative approach recommended previously (Delpont et al., 2010) was employed in ascertaining the positive or negative selection at each site. The sites which were found to be non-neutral and be positively or negatively selected by at least three methods at significant levels were considered to be positively/negatively selected sites. The p-values = < 0.1 were considered as significant for SLAC, FEL and MEME methods, whereas the Bayes factor  $\geq 30$  was used as the significance threshold for FUBAR method.

## **Results And Discussion**

### **Cosmopolitan genotype demonstrated inter-continental dispersal through an extensive spatial network**

A spatial analysis of DENV-2 sequences of all known genotypes was conducted to determine their distribution patterns. All complete *E* gene sequences (1,485 bp) of DENV-2 as of July 2019 were retrieved from the NCBI database and tagged with the country of origin (location) and year of isolation (time) information available in NCBI flat files. The final alignment included 4,551 sequences. Each sequence was classified into a known genotype, based on the phylogenetic relationship. Sequences of each genotype were mapped to determine the global distribution (Figure 1). The analysis demonstrated a variable degree of geo-distribution of each genotype. Cosmopolitan genotype was the most widely distributed, though was remarkably absent in the Americas. Asian II and Asian-American genotypes were present in Southeast Asia as well as Central and South America. American genotype was the least

widespread. Asian I genotype was particularly restricted to Southeast Asia. Southeast Asia showed the highest genotype heterogeneity and was the most notable global epicenter of DENV-2 transmission.

Bayesian Stochastic Search Variable studies (BSSVS)-based phylogeography analyses were conducted to determine the spatio-temporal dispersal and the most active hubs of cosmopolitan genotype (Figure 2). The analyses showed an extensive dispersal network of short- and long-distance links. All BSSVS links achieved high statistical support (Bayes Factor >3 and posterior probability  $\geq 0.75$ ). Inter-continental dispersal was notable among the long-distance links.

The BSSVS findings reflected the epidemiological reality of cosmopolitan genotype's presence in a wide geographical scale [20, 33, 52-59]. The most intense virus activity was visible in Southeast Asia, mainly in two epicenters; first, an area covering Singapore, Malaysia and Indonesia and the second centered on Taiwan and China in the Far East. Both epicenters actively exchanged virus strains with Australasia and the Pacific regions. The Far East hub also demonstrated long-range dispersal links with East Africa. However, both hubs were not linked to the Middle East. Instead, Middle East demonstrated alternative dispersal links with the Indian sub-continent. Indian sub-continent also showed decisive links with Southeast Asian and Far-East epicenters, suggesting active virus exchange between the sub-continent and other regions. The extensive network demonstrated how the global trade and travel have facilitated the geo expansion of DENV-2 cosmopolitan genotype. The long-range dispersal of DENV is possible either through the migration of humans, infected mosquitoes or desiccated mosquito eggs. Human migration is unarguably the more likely mode of distant virus dispersal than live mosquito transfer. Virus dispersal through desiccated mosquito eggs is less plausible than other modes due to limited evidence of vertical transmission of DENV [60]. The rapid expansion of air travel in recent decades has immensely facilitated human migration and thereby DENV dispersal [61]. For instance, Singapore is one of the most connected global aviation hubs and showed numerous inbound and outbound dispersal links in our BSSVS analysis (Figure 1). China's economic and trade expansion resonates well with its long-distance outbound virus dispersal to the African continent. The substantial dispersal of DENV-2 from Indian sub-continent to the Middle East might be due to increased trade and pilgrimage travel [62].

Overall the extensive dispersal of cosmopolitan genotype suggests its ability to adapt to vector and host populations in a wide geographical range. This contrasts with DENV-2 genotypes such as American and Asian I, which are largely restricted to certain geographies [5]. However, cosmopolitan genotype seemed to have struggled to establish in locations where these geographically restricted genotypes dominated. For example, of 4,551 *E* gene sequences analyses in the present study, Asian I genotype (70.3%, n=333) was the most common among 474 sequences from Vietnam. This was in comparison to 7.6% (n=33) of cosmopolitan genotype sequences. The same pattern was observed among 322 sequences from Thailand (91.3% Asian I; n=294 and 5.9% cosmopolitan; n=19). Therefore, the ability of cosmopolitan genotype to survive in different environments seemed to be variable. One of the contributory factors for this variability might be evolutionary adaptations that generate heterogeneity, confer fitness advantages, and cause selective sweeps, allowing virus lineages to dominate differentially in various geographies

[63]. This warranted further analyses of evolutionary signatures, heterogeneity, sub-lineage geo-clustering and natural selection of the cosmopolitan genotype.

### **Cosmopolitan genotype demonstrated high intra-genotype heterogeneity**

The Bayesian phylogeny of DENV-2 complete coding sequences revealed distinct clustering of each genotype, with strong posterior probability support (Figure 3). The cosmopolitan genotype comprised of two distinct groups. One group included sequences belonging to the Indian sub-continent lineage, which has largely been reported in the Indian sub-continent [64, 65]. The other group included the rest of cosmopolitan genotype sequences reported from other countries, especially in Southeast Asia.

The mean nucleotide substitution rate of each genotype ranged from  $10^{-4}$  to  $10^{-3}$  subs/site/year (Table 1). This estimate agreed with the empirical data [62, 64, 66], demonstrating comparable evolutionary rates among all genotypes of DENV-2, regardless of their differences in distribution. The inferred mean root age of DENV-2 was  $\sim 287.32$  years from 2019 (95% highest posterior density (HPD) - 164.77 to 440.30 years), which was comparable to previous estimates of 325.2 years [62]. The tMRCA data also suggested that cosmopolitan genotype might have emerged in parallel to Asian I and II genotypes between 1944 and 1964. The analysis predicted a major split in cosmopolitan genotype, with the emergence of Indian sub-continent lineage in early 1980s.

### **Table 1. Time to most recent common ancestor (tMRCA) analysis of DENV-2 genotypes**

Description	Node height (95% HPD) in years	Estimated year of emergence <sup>¶</sup>
Root ancestor	287.32 (164.77- 440.3)	1745
Sylvatic	140 (87.9-208.46)	1886
Cosmopolitan genotype (other lineages)	64.45 (54.96- 75.05)	1955
Cosmopolitan genotype (Indian sub- continent lineage)	36.1 (30.6-41.92)	1983
Asian II genotype	74.24 (64.14- 86.89)	1939
Asian I genotype	62 (49.06-77.62)	1946
American genotype	125.26 (90.46- 174.85)	1899
Asian-American genotype	51.12 (43.78- 60.13)	1968

<sup>¶</sup>Estimated year of emergence was calculated by subtracting the node height from 2019, which was the latest temporal data point in the analysis. HPD=Highest posterior density

The cosmopolitan genotype was distinct from other genotypes by a 12-amino acid signature (Table 2). Among non-sylvatic genotypes of DENV-2, each residue of this signature was “fixed” and unique to the cosmopolitan genotype. The signature mutations were spread across the structural and non-structural proteins (PreM/M, Envelope, NS3, NS4B and NS5), implying a genome-wide selection of amino acid residues during the evolutionary history of cosmopolitan genotype. Whether these residues conferred a fitness advantage to cosmopolitan genotype warrant further investigations.

**Table 2. Unique amino acid signature of cosmopolitan genotype.**

Polyprotein position	Gene	Gene position	REF	Other genotypes	Cosmopolitan genotype (% of fixation <sup>¶</sup> )
166	preM/M	52	K	K	N (100%)
196	preM/M	82	T	T	A (99%)
262	preM/M	148	H	H	Y (96%)
351	E	71	E/D	E/D	A (100%)
429	E	149	H	H	N (99%)
670	E	390	N/D	N/D	S (97%)
742	E	462	I	I	V (99%)
1635	NS3	160	A/S	A/S	T (100%)
2256	NS4B	13	L	L	F (99%)
2496	NS5	5	I/V/M	I/V/M	T (99%)
2506	NS5	15	S	S	N (100%)
2826	NS5	335	V/L	V/L	I (100%)

<sup>¶</sup>The percentage values indicate the extent of fixation of each mutation in 445 sequences of cosmopolitan genotype as compared to 1,007 sequences of other genotypes of DENV-2.

M=membrane; NS=non-structural proteins; preM=pre-membrane

A separate phylogenetic analysis was performed by using complete *E* gene sequences to determine the intra-genotype heterogeneity of cosmopolitan genotype. Here, *E* gene was preferred to complete genomes because of the availability of more *E* gene sequences from different geographical regions. *E* gene is also a widely used candidate for the phylogenetic classification of DENV lineages [13, 67, 68]. The analysis included 2,392 complete *E* gene sequences of cosmopolitan genotype reported from 39 countries (Supplementary Table 2).

Similar to the whole genome-based phylogeny, *E* gene tree also branched into two main groups at the root, separating the Indian sub-continent lineage from the rest of cosmopolitan genotype (other lineages) sequences (Figure 4). Besides, other lineage sequences clustered into five distinct lineages, suggesting an overall high heterogeneity of the cosmopolitan genotype (Figure 4). Based on the nucleotide and amino acid similarity matrices (Table 3), Indian sub-continent lineage was the most divergent among all lineages within the cosmopolitan genotype.

**Table 3. *E* gene nucleotide similarity matrix between different lineages of the cosmopolitan genotype**

Other or Indian sub-continent lineages	Other lineages					Indian sub-continent lineage
	Lineage-1	Lineage-2	Lineage-3	Lineage-4	Lineage-5	
Lineage-1	1.00	0.97	0.97	0.97	0.97	0.94
Lineage-2	0.97	1.00	0.97	0.97	0.96	0.93
Lineage-3	0.97	0.97	1.00	0.97	0.97	0.93
Lineage-4	0.97	0.97	0.97	1.00	0.98	0.94
Lineage-5	0.97	0.96	0.97	0.98	1.00	0.94
Indian sub-continent lineage	0.94	0.93	0.93	0.94	0.94	1.00

The geographical distribution of these lineages varied within the complete dataset. Notably, sequences from Indonesia positioned at the base of all lineages, except the Indian sub-continent lineage. On the other hand, sequences from Singapore, Malaysia and China appeared in all lineages. While all lineages were spread across different countries, certain lineages dominated in a particular country or a region. For example, lineage 1 almost entirely comprised sequences from the Philippines. There were two outgroups of lineage 1 that included sequences from the Pacific Islands (founder outgroup) and Indonesia (secondary outgroup), suggesting a likelihood of the earliest ancestry of lineage 1 either in the Pacific Islands or Indonesia, and subsequent introduction into the Philippines. Interestingly, the phylogeography analysis showed a decisive direct link between the Philippines and the Pacific, and an indirect decisive link with Indonesia, through Papua New Guinea (Figure 2). Root state posterior probability data extracted from the MCC tree suggested that the most likely common ancestor of lineage 1 originated in Indonesia. This strain might have spread to the Philippines via Papua New Guinea as shown in the phylogeography analysis (Figure 2). A similar observation was also made in the Indian sub-continent lineage that is largely circulating in the Indian sub-continent countries. The earliest available sequence in the basal clade of this lineage (outgroup) was a sequence reported from India in 1974 (NCBI Accession FJ538920). Two sub-lineages (sub-lineage 1 and 2) emerged from this outgroup in early 1990s. One of them showed limited evidence of outbound spread from the Indian sub-continent, mainly to the Middle East and East Africa. In contrast, the other sub-lineage showed relatively wide dispersal from the Middle East to Southeast and East Asia. The available data showed records of both sub-lineages as late as 2017-18.

### **Is Indian sub-continent lineage becoming a new genotype?**

Having been genetically and geographically distinct from the rest of cosmopolitan lineages (Figure 4), we investigated whether the Indian sub-continent lineage is divergent enough to be classified as a new genotype. To achieve this, we analyzed the genetic divergence of Indian sub-continent lineage from other known non-sylvatic genotypes (Asian I, Asian II, Asian-American and American genotypes). These analyses included complete *E* gene sequences of all cosmopolitan genotype sequences used in the current study, but split into two categories, namely “Indian sub-continent lineage” and “other lineages”.

Other lineages category included lineages 1-5 illustrated in Figure 4. The dataset also included sequences of Asian I, Asian II, Asian-American and American genotypes that were selected to represent the genetic and geographical variations within each genotype. First, *E* gene nucleotide and amino acid similarity scores were calculated and compared between the Indian sub-continent and other lineages of cosmopolitan genotype and other genotypes (Table 4).

**Table 4. Comparison of *E* gene divergence matrices between the Indian sub-continent lineage and other lineages of cosmopolitan genotype and other genotypes of DENV-2.**

Genotype	Nucleotide dissimilarity (Amino acid dissimilarity)	
	Cosmopolitan (Other lineages)	Cosmopolitan (Indian sub-continent lineage)
Cosmopolitan (other lineages)	0% (0%)	5.6% (5.3%)
Cosmopolitan (Indian sub-continent lineage)	5.6% (5.3%)	0% (0%)
Asian II	8.8% (7.9%)	8.6% (7.4%)
Asian I	8.8% (8.8%)	8.9% (8.2%)
American	10.2% (10%)	9.4% (9%)
Asian/American	7.8% (8.5%)	7.1% (7.7%)

The nucleotide dissimilarity between the cosmopolitan (other lineages) and Indian sub-continent lineage was 5.6% (Table 4). Respective values between the cosmopolitan (other lineages) and other genotypes ranged between 7.8% (Asian-American) and 10.2% (American). The Indian sub-continent lineage also demonstrated comparable nucleotide divergence with other genotypes of DENV-2 (divergence range: 7.1%-9.4%). Empirical data has suggested a nucleotide divergence (range 3-6%) threshold of 6% within *E* gene among different genotypes of DENV-2 [7, 14, 69]. Our data showed that nucleotide divergence between the cosmopolitan (other lineages) and Indian sub-continent lineage (5.6%) is on the verge of this threshold. To further corroborate the genetic distinction between Indian sub-continent and other lineages of cosmopolitan genotype, we compared  $\pi$  nucleotide diversity index values of *E* gene sequences of cosmopolitan genotype (split into Indian sub-continent and other lineages) and other genotypes of DENV-2. As shown in Figure 5, cosmopolitan genotype demonstrated the highest nucleotide diversity among all non-sylvatic genotypes of DENV-2. Moreover, Indian sub-continent and other lineages formed distinct subpopulations within the cosmopolitan genotype, with ~6% *E* gene pairwise distance between the two groups. The  $\pi$  index of Indian sub-continent lineage overlapped with a batch of Asian II genotype sequences, further corroborating the genetic distinction of Indian sub-continent lineage from other lineages of cosmopolitan genotype. These observations suggested that Indian sub-continent lineage has

already diverged from the rest of cosmopolitan genotype lineages and is almost at the threshold of 6% nucleotide distance within *E* gene to qualify as a separate genotype.

To further support the genetic distinction of Indian sub-continent lineage within cosmopolitan genotype, we compared complete coding sequences of Indian sub-continent and other lineages. The analysis showed a 16-amino acid signature unique to the Indian sub-continent lineage (Figure 6, Supplementary Table 3).

Eight amino acid residues of this signature were also shared by a small subset of Indonesian sequences of the cosmopolitan genotype, isolated during 1974-75 (Table 5). Based on NCBI data (latest by July 2019), the earliest sequence of Indian sub-continent lineage is a strain (NIV\_742295) isolated in 1974 (NCBI Accession FJ538920). NCBI Accession FJ538920 is a partial genome sequence and was thus not included in the whole genome based tMRCA analysis that estimated the origin of Indian sub-continent lineage in early 1980s (Table 1). Therefore, the partial genome data suggested that Indian sub-continent lineage might have emerged even earlier than early 1980s. This was corroborated by the group of Indonesian sequences that were reported at the same time as the earliest genetic record of Indian sub-continent lineage in 1974. Sharing of amino acid residues unique to Indian sub-continent lineage by 1974-75 Indonesian sequences also implied a common ancestry between the two groups.

**Table 5. Unique amino acid substitutions shared between the Indian sub-continent lineage and Indonesian sequences (1974-76).**

Polyprotein position	Gene	Gene position	REF	Cosmopolitan genotype		
				Indonesian (1974-76; n=07)	Indian sub-continent lineage (n=94)	Other lineages (n=344)
143	preM	29	D	D (86%), N (14%)	D (100%)	N (100%)
332	E	52	Q	Q (71%), H (29%)	Q (100%)	H (97%)
906	NS1	131	Q	Q (71%), H (29%)	Q (100%)	H (98%)
1408	NS2B	63	D	D (100%)	D (100%)	E (98%)
1748	NS3	273	V	V (100%)	V (100%)	I (98%)
1936	NS3	461	I	I (86%), V (14%)	I (98%)	V (97%)
2589	NS5	98	R	R (71%), K (29%)	R (98%)	K (98%)
3096	NS5	605	G	G (71%), V (29%)	G (100%)	V (98%)

Percentage values indicate the extent of fixation of each mutation in respective groups of sequences. NS=non-structural proteins; preM=pre-membrane

### **Selection pressures acted on different amino acid residues between Indian sub-continent and other lineages of cosmopolitan genotype**

The spatio-temporal dominance of virus genotypes is influenced by selective sweeps that partly occur during virus-host adaptations. In arboviruses, such as DENV-2, these adaptations occur in both vertebrate and invertebrate hosts [70]. Therefore, we hypothesized that selection pressure analyses among different lineages would depict amino acid residues that altered during evolution and dispersal of the cosmopolitan genotype. To achieve this, differential rates of nonsynonymous and synonymous substitution rates per site were estimated in the complete coding sequences of cosmopolitan lineages (447 sequences of other lineages and 98 sequences of Indian sub-continent lineage) using SLAC, FEL, FUBAR and MEME methods [71-74]. The findings showed that majority of residues were neutral and purifying selection was more pronounced than the positive selection. This was in agreement with previous studies [71, 75]. Among the few positively-selected sites, there was no overlap between the

Indian sub-continent lineage and other lineages, suggesting differences in selective pressures acting on these lineages. NS2-1160 and NS5-3122 residues showed evidence of positive selection in the Indian sub-continent lineage. Valine was present at residue NS2-1160 in 77% of Indian sub-continent lineage sequences, whereas Isoleucine was present at the same position of other lineages. On the other hand, the positive selection was evident on C-9R, NS2-1298A, NS3-1867S and NS4-3139V residues in other lineages of the cosmopolitan genotype (Table 6). At residue NS5-3122, 35% of Indian sub-continent lineage sequences possessed Glycine. In contrast, Serine was present at the same residue in ~99% of sequences of other lineages.

**Table 6. Selection pressure analysis of the complete coding region of cosmopolitan genotype**

Lineage	Protein	Amino acid position	SLAC	FEL	FUBAR	MEME
Lineages 1-5	C	<b>9</b>	<b>0.05</b>	<b>0.03</b>	<b>89.45</b>	0.67
	C	95	0.14	0.09	46.73	0.11
	E	627	0.26	0.22	29.06	0.07
	E	<b>670</b>	0.11	<b>0.02</b>	<b>214.78</b>	0.67
	NS1	953	0.18	0.16	<b>48.62</b>	0.16
	NS2	<b>1298</b>	<b>0.02</b>	<b>0.03</b>	<b>88.23</b>	<b>0.04</b>
	NS3	<b>1867</b>	<b>0.04</b>	<b>0.03</b>	<b>134.29</b>	<b>0.02</b>
Indian sub-continent lineage	NS5	<b>3139</b>	<b>0.03</b>	0.23	<b>80.97</b>	0.14
	C	130	0.34	0.18	9.69	0.2
	C	171	0.41	0.19	10.86	0.1
	E	332	0.28	0.14	14.86	0
	E	640	0.29	0.14	20.05	0.67
	NS1	967	0.34	0.19	9.61	0.01
	NS2	<b>1160</b>	<b>0.07</b>	<b>0.06</b>	<b>65.28</b>	0.67
NS5	<b>3122</b>	0.16	<b>0.04</b>	<b>91.78</b>	0.67	

Positive selection was evident on residues in bold font, based on the selection pressure analyses. The values shown for each method are p-values, except for the FUBAR method for which the values indicate the Bayes factor. The p-values = < 0.1 were considered as significant for SLAC, FEL, and MEME methods, whereas the Bayes factor >=30 was used for FUBAR method. C=Capsid; E=envelope; NS=non-structural proteins

## Conclusions

The current study demonstrated that cosmopolitan genotype has spread through an extensive spatial network, allowing distinct intra-genotype lineages to dominate in different geographies. High intra-genotype heterogeneity reflected intense evolutionary forces acting within the cosmopolitan genotype. Findings indicated a major split in the cosmopolitan genotype, and suggested that Indian sub-continent lineage is maturing into a separate genotype. The notable genetic characteristics that appeared during the evolution and global expansion of cosmopolitan genotype warrant further investigations into the phenotype differences, host adaptation and survival of its lineages that affect DENV-2 epidemiology in endemic settings.

## Abbreviations

BF: Bayes factor; BSSVS: Bayesian Stochastic Search Variable studies; DENV: Dengue virus; E: Envelope gene; ESS: effective sample size; FEL: fixed effects likelihood; FUBAR: fast unbiased Bayesian approximation; GTR: General time reversible; HPD: highest posterior density; M=membrane; MCC: maximum clade credibility; MCMC: Markov Chain Monte Carlo; MEME: mixed effect model of evolution; NS=non-structural proteins; preM=pre-membrane; SLAC: Single likelihood ancestor counting; tMRCA: time to most recent common ancestor

## Declarations

### Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Declaration of Competing Interest

Authors declare no competing financial or non-financial interests.

### Acknowledgements

We thank general practitioners in Singapore for providing clinical samples from dengue-suspected patients. We are also grateful to colleagues from Diagnostics and Applied Entomology branches at the Environmental Health Institute, Singapore for their support in providing dengue-positive sera and mosquito specimens. We also thank Dr. Judith Wong for her valuable inputs and guidance. The computational work for this article was partially performed on resources of the National Supercomputing Centre, Singapore (<https://www.nscg.sg>) under the project ID 14001593. This study was supported by the

National Environment Agency, Singapore, which had no role in the study design, data collection, data analysis, data interpretation, writing of the report, or in the decision to submit the paper for publication.

## Author Contributions Statement

Surya Pavan Yenamandra: Data curation, formal analysis, writing - original draft preparation  
Carmen Koo: Resources, data curation, formal analysis  
Suzanna Chiang: Resources, data curation  
Lim Han Shi Jeri: Data curation, formal analysis  
Zhen Yuan Yeo: Data curation, formal analysis  
Lee Ching Ng: Conceptualization, writing – reviewing & editing  
Hapuarachchige Chanditha Hapuarachchi: Conceptualization, supervision, Methodology, writing - original draft preparation.

## Data availability

Envelope gene sequences and complete genome sequences generated during the present study were deposited in GenBank database under the accession numbers MW510024 - MW510156, MW510238 - MW510994 and MW512341 - MW512498.

## References

1. WHO, *Global strategy for dengue prevention and control 2012-2020*. World Health Organization. . 2012.
2. Stanaway, J.D., et al., *The global burden of dengue: an analysis from the Global Burden of Disease Study 2013*. *Lancet Infect Dis*, 2016. **16**(6): p. 712-723.
3. Messina, J.P., et al., *Global spread of dengue virus types: mapping the 70 year history*. *Trends Microbiol*, 2014. **22**(3): p. 138-46.
4. Putri, D.H., et al., *Immunogenicity of a Candidate DNA Vaccine Based on the prM/E Genes of a Dengue Type 2 Virus Cosmopolitan Genotype Strain*. *Jpn J Infect Dis*, 2015. **68**(5): p. 357-63.
5. Chen, R. and N. Vasilakis, *Dengue—quo tu et quo vadis?* *Viruses*, 2011. **3**(9): p. 1562-608.
6. Rico-Hesse, R., *Molecular evolution and distribution of dengue viruses type 1 and 2 in nature*. *Virology*, 1990. **174**(2): p. 479-93.
7. Holmes, E.C. and S.S. Twiddy, *The origin, emergence and evolutionary genetics of dengue virus*. *Infect Genet Evol*, 2003. **3**(1): p. 19-28.
8. Cologna, R., P.M. Armstrong, and R. Rico-Hesse, *Selection for virulent dengue viruses occurs in humans and mosquitoes*. *J Virol*, 2005. **79**(2): p. 853-9.
9. Cologna, R. and R. Rico-Hesse, *American genotype structures decrease dengue virus output from human monocytes and dendritic cells*. *J Virol*, 2003. **77**(7): p. 3929-38.
10. Guo, C., et al., *Global Epidemiology of Dengue Outbreaks in 1990-2015: A Systematic Review and Meta-Analysis*. *Front Cell Infect Microbiol*, 2017. **7**: p. 317.
11. Aguas, R., et al., *Cross-serotype interactions and disease outcome prediction of dengue infections in Vietnam*. *Sci Rep*, 2019. **9**(1): p. 9395.

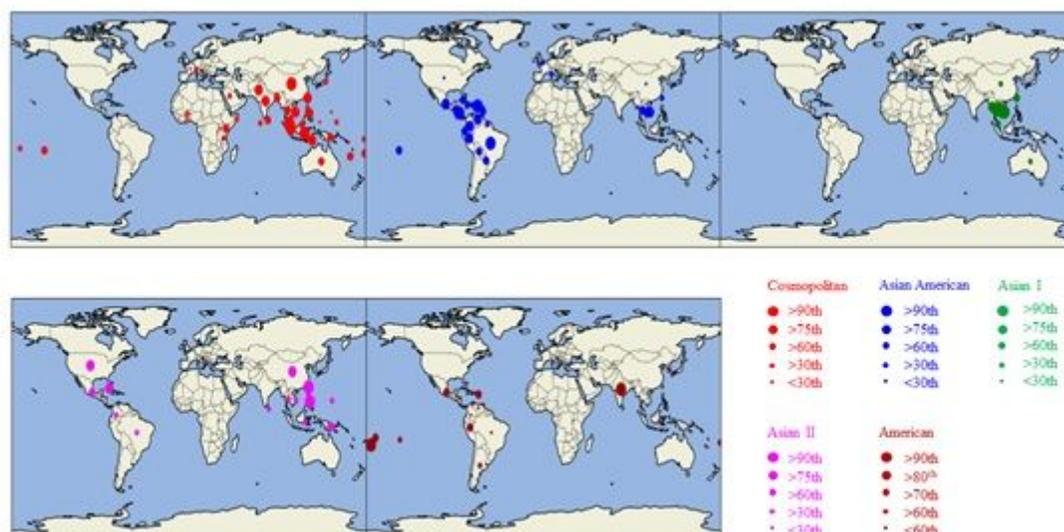
12. Fried, J.R., et al., *Serotype-specific differences in the risk of dengue hemorrhagic fever: an analysis of data collected in Bangkok, Thailand from 1994 to 2006*. PLoS Negl Trop Dis, 2010. **4**(3): p. e617.
13. Twiddy, S.S., et al., *Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus*. Virology, 2002. **298**(1): p. 63-72.
14. Weaver, S.C. and N. Vasilakis, *Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease*. Infect Genet Evol, 2009. **9**(4): p. 523-40.
15. Hill, S.C., et al., *Early Genomic Detection of Cosmopolitan Genotype of Dengue Virus Serotype 2, Angola, 2018*. Emerg Infect Dis, 2019. **25**(4): p. 784-787.
16. Khan, M.A., et al., *Emergence and diversification of dengue 2 cosmopolitan genotype in Pakistan, 2011*. PLoS One, 2013. **8**(3): p. e56391.
17. Lustig, Y., et al., *An outbreak of dengue virus (DENV) type 2 Cosmopolitan genotype in Israeli travellers returning from the Seychelles, April 2017*. Euro Surveill, 2017. **22**(26).
18. Phadungsombat, J., et al., *Emergence of genotype Cosmopolitan of dengue virus type 2 and genotype III of dengue virus type 3 in Thailand*. PLoS One, 2018. **13**(11): p. e0207220.
19. Soh, L.T., et al., *External quality assessment of dengue and chikungunya diagnostics in the Asia Pacific region, 2015*. Western Pac Surveill Response J, 2016. **7**(2): p. 26-34.
20. Banu, S., et al., *Dengue transmission in the Asia-Pacific region: impact of climate change and socio-environmental factors*. Trop Med Int Health, 2011. **16**(5): p. 598-607.
21. Kuno, G., *Research on dengue and dengue-like illness in East Asia and the Western Pacific during the First Half of the 20th century*. Rev Med Virol, 2007. **17**(5): p. 327-41.
22. Cantelar de Francisco, N., *[Dengue hemorrhagic fever. Southeast Asia and the Western Pacific (review article) (I)]*. Rev Cubana Med Trop, 1983. **35**(2): p. 130-5.
23. Ayolabi, C.I., et al., *Detection of Dengue viruses among febrile patients in Lagos, Nigeria and phylogenetics of circulating Dengue serotypes in Africa*. Infect Genet Evol, 2019. **75**: p. 103947.
24. Yamamoto, S.P., et al., *Dengue Virus in Traveler Returning to Japan from the Democratic Republic of the Congo, 2015*. Jpn J Infect Dis, 2019. **72**(6): p. 426-428.
25. Okada, K., et al., *Dengue Virus Type 1 Infection in Traveler Returning from Tanzania to Japan, 2019*. Emerg Infect Dis, 2019. **25**(9): p. 1782-1784.
26. Matsui, T., et al., *Dengue Virus Type 2 Infection in a Traveler Returning from Saudi Arabia to Japan*. Jpn J Infect Dis, 2019. **72**(5): p. 340-342.
27. Tsuboi, M., et al., *Dengue Virus Type 2 in Travelers Returning to Japan from Sri Lanka, 2017*. Emerg Infect Dis, 2017. **23**(11).
28. Suzuki, T., et al., *Dengue Virus Exported from Cote d'Ivoire to Japan, June 2017*. Emerg Infect Dis, 2017. **23**(10).
29. Jeong, Y.E., et al., *Comparison of the Epidemiological Aspects of Imported Dengue Cases between Korea and Japan, 2006-2010*. Osong Public Health Res Perspect, 2016. **7**(1): p. 71-4.

30. Tsheten, T., et al., *Spatial and temporal patterns of dengue incidence in Bhutan: a Bayesian analysis*. *Emerg Microbes Infect*, 2020. **9**(1): p. 1360-1371.
31. Zangmo, S., et al., *Molecular epidemiology of dengue fever outbreaks in Bhutan, 2016-2017*. *PLoS Negl Trop Dis*, 2020. **14**(4): p. e0008165.
32. Zangmo, S., et al., *Epidemiological and Molecular Characterization of Dengue Virus Circulating in Bhutan, 2013-2014*. *PLoS Negl Trop Dis*, 2015. **9**(8): p. e0004010.
33. Dorji, T., et al., *Diversity and origin of dengue virus serotypes 1, 2, and 3, Bhutan*. *Emerg Infect Dis*, 2009. **15**(10): p. 1630-2.
34. Hapuarachchi, H.C., et al., *Epidemic resurgence of dengue fever in Singapore in 2013-2014: A virological and entomological perspective*. *BMC Infect Dis*, 2016. **16**: p. 300.
35. Rajarethinam, J., et al., *Dengue in Singapore from 2004 to 2016: Cyclical Epidemic Patterns Dominated by Serotypes 1 and 2*. *Am J Trop Med Hyg*, 2018. **99**(1): p. 204-210.
36. Nakamura, T., et al., *Parallelization of MAFFT for large-scale multiple sequence alignments*. *Bioinformatics*, 2018. **34**(14): p. 2490-2492.
37. Hapuarachchi, H.C., et al., *Intra-epidemic evolutionary dynamics of a Dengue virus type 1 population reveal mutant spectra that correlate with disease transmission*. *Sci Rep*, 2016. **6**: p. 22592.
38. Tan, C.H., et al., *Evaluation of the Dengue NS1 Ag Strip(R) for detection of dengue virus antigen in Aedes aegypti (Diptera: Culicidae)*. *Vector Borne Zoonotic Dis*, 2011. **11**(6): p. 789-92.
39. Kumar, S., G. Stecher, and K. Tamura, *MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets*. *Mol Biol Evol*, 2016. **33**(7): p. 1870-4.
40. Bouckaert, R., et al., *BEAST 2: a software platform for Bayesian evolutionary analysis*. *PLoS Comput Biol*, 2014. **10**(4): p. e1003537.
41. Drummond, A.J., et al., *Bayesian phylogenetics with BEAUti and the BEAST 1.7*. *Mol Biol Evol*, 2012. **29**(8): p. 1969-73.
42. Darriba, D., et al., *jModelTest 2: more models, new heuristics and parallel computing*. *Nat Methods*, 2012. **9**(8): p. 772.
43. Lemey, P., et al., *Bayesian phylogeography finds its roots*. *PLoS Comput Biol*, 2009. **5**(9): p. e1000520.
44. Drummond, A.J., et al., *Relaxed phylogenetics and dating with confidence*. *PLoS Biol*, 2006. **4**(5): p. e88.
45. Bielejec, F., et al., *SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics*. *Bioinformatics*, 2011. **27**(20): p. 2910-2912.
46. Faria, N.R., et al., *Simultaneously reconstructing viral cross-species transmission history and identifying the underlying constraints*. *Philos Trans R Soc Lond B Biol Sci*, 2013. **368**(1614): p. 20120196.
47. Waterhouse, A.M., et al., *Jalview Version 2—a multiple sequence alignment editor and analysis workbench*. *Bioinformatics*, 2009. **25**(9): p. 1189-91.

48. Hall, T.A., *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT*. Nucl. Acids. Symp. Ser, 1999. **41**: p. 95-98.
49. Nei, M. and W.H. Li, *Mathematical model for studying genetic variation in terms of restriction endonucleases*. Proc Natl Acad Sci U S A, 1979. **76**(10): p. 5269-73.
50. Cock, P.J., et al., *Biopython: freely available Python tools for computational molecular biology and bioinformatics*. Bioinformatics, 2009. **25**(11): p. 1422-3.
51. Sukumaran, J. and M.T. Holder, *DendroPy: a Python library for phylogenetic computing*. Bioinformatics, 2010. **26**(12): p. 1569-71.
52. Dinu, S., et al., *Molecular epidemiology of dengue fever cases imported into Romania between 2008 and 2013*. Travel Med Infect Dis, 2015. **13**(1): p. 69-73.
53. El-Kafrawy, S.A., et al., *Multiple Introductions of Dengue 2 Virus Strains into Saudi Arabia from 1992 to 2014*. Vector Borne Zoonotic Dis, 2016. **16**(6): p. 391-9.
54. Ernst, T., et al., *Emergence of a new lineage of dengue virus type 2 identified in travelers entering Western Australia from Indonesia, 2010-2012*. PLoS Negl Trop Dis, 2015. **9**(1): p. e0003442.
55. Gathii, K., et al., *Complete Coding Sequences of Dengue Virus Type 2 Strains from Febrile Patients Seen in Malindi District Hospital, Kenya, during the 2017 Dengue Fever Outbreak*. Genome Announc, 2018. **6**(15).
56. Gupta, B.P., et al., *Re-emergence of dengue virus serotype 2 strains in the 2013 outbreak in Nepal*. Indian J Med Res, 2015. **142 Suppl**: p. S1-6.
57. Hamel, R., et al., *Phylogenetic analysis revealed the co-circulation of four dengue virus serotypes in Southern Thailand*. PLoS One, 2019. **14**(8): p. e0221179.
58. Hamid, Z., et al., *Molecular Investigation of Dengue Virus Serotype 2 Circulation in Kassala State, Sudan*. Jpn J Infect Dis, 2019. **72**(1): p. 58-61.
59. Langat, S.K., et al., *Origin and evolution of dengue virus type 2 causing outbreaks in Kenya: Evidence of circulation of two cosmopolitan genotype lineages*. Virus Evol, 2020. **6**(1): p. veaa026.
60. Grunnill, M. and M. Boots, *How Important is Vertical Transmission of Dengue Viruses by Mosquitoes (Diptera: Culicidae)?* J Med Entomol, 2016. **53**(1): p. 1-19.
61. Gubler, D.J., *Dengue, Urbanization and Globalization: The Unholy Trinity of the 21(st) Century*. Trop Med Health, 2011. **39**(4 Suppl): p. 3-11.
62. Walimbe, A.M., et al., *Global phylogeography of Dengue type 1 and 2 viruses reveals the role of India*. Infect Genet Evol, 2014. **22**: p. 30-9.
63. Stephan, W., *Selective Sweeps*. Genetics, 2019. **211**(1): p. 5-13.
64. Afreen, N., et al., *Evolutionary Analysis of Dengue Serotype 2 Viruses Using Phylogenetic and Bayesian Methods from New Delhi, India*. PLoS Negl Trop Dis, 2016. **10**(3): p. e0004511.
65. Ujwal, S., et al., *Circulation of Asian-I and Cosmopolitan genotypes of Dengue-2 virus in northeast India, 2016-2017*. J Vector Borne Dis, 2019. **56**(3): p. 231-236.

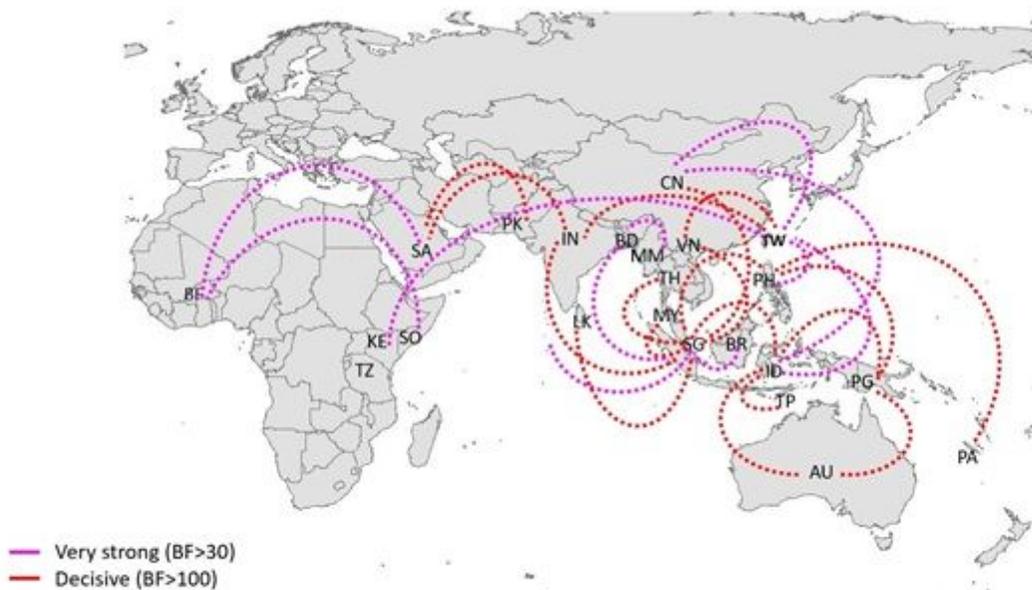
66. Wei, K. and Y. Li, *Global evolutionary history and spatio-temporal dynamics of dengue virus type 2*. Sci Rep, 2017. **7**: p. 45505.
67. Marin, M.S., et al., *Phylogeny of TYU, SRE, and CFA virus: different evolutionary rates in the genus Flavivirus*. Virology, 1995. **206**(2): p. 1133-9.
68. Zanotto, P.M., et al., *Population dynamics of flaviviruses revealed by molecular phylogenies*. Proc Natl Acad Sci U S A, 1996. **93**(2): p. 548-53.
69. Parameswaran, P., et al., *Genome-wide patterns of intrahuman dengue virus diversity reveal associations with viral phylogenetic clade and interhost diversity*. J Virol, 2012. **86**(16): p. 8546-58.
70. Lequime, S., et al., *Genetic Drift, Purifying Selection and Vector Genotype Shape Dengue Virus Intra-host Genetic Diversity in Mosquitoes*. PLoS Genet, 2016. **12**(6): p. e1006111.
71. Yang, Z., et al., *Codon-substitution models for heterogeneous selection pressure at amino acid sites*. Genetics, 2000. **155**(1): p. 431-49.
72. Kosakovsky Pond, S.L. and S.D. Frost, *Not so different after all: a comparison of methods for detecting amino acid sites under selection*. Mol Biol Evol, 2005. **22**(5): p. 1208-22.
73. Murrell, B., et al., *FUBAR: a fast, unconstrained bayesian approximation for inferring selection*. Mol Biol Evol, 2013. **30**(5): p. 1196-205.
74. Murrell, B., et al., *Detecting individual sites subject to episodic diversifying selection*. PLoS Genet, 2012. **8**(7): p. e1002764.
75. Anez, G., M.E. Morales-Betoulle, and M. Rios, *Circulation of different lineages of dengue virus type 2 in Central America, their evolutionary time-scale and selection pressure analysis*. PLoS One, 2011. **6**(11): p. e27459.

## Figures



## Figure 1

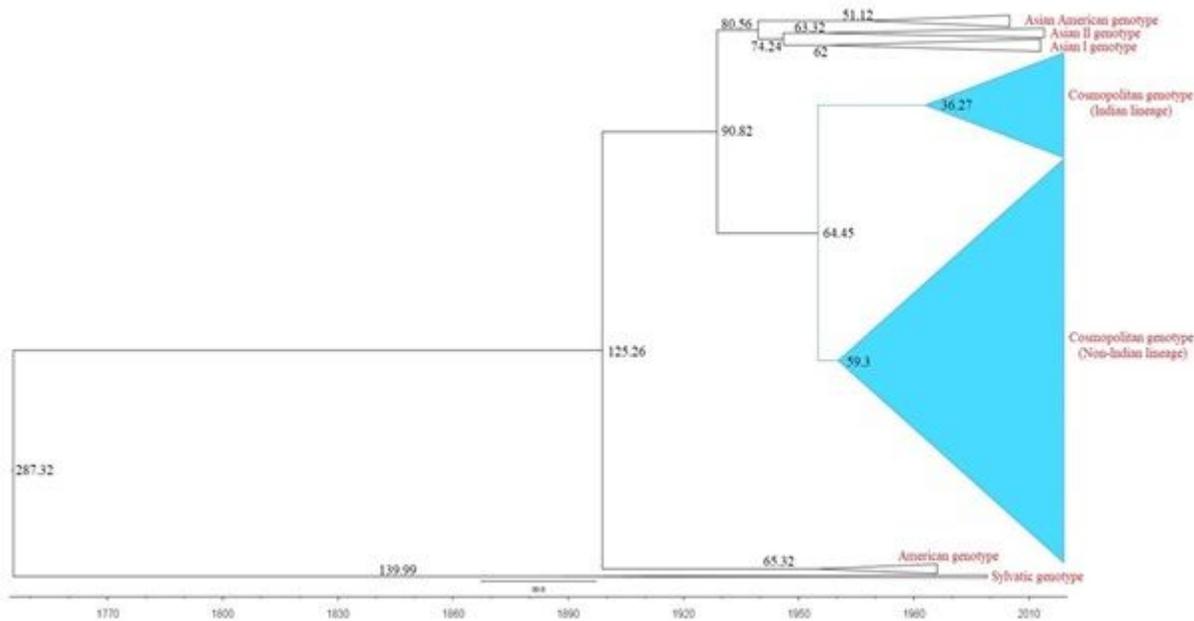
Spatial distribution of DENV-2 genotypes. Genotypes are represented by different colors as shown in the legend. The number of sequences belonging to each genotype reported by all countries was ranked into >90th, 75-90th, 60-74th, 30-59th and <30th percentile categories that are shown in circles of different size. Due to low sample size, the percentile ranking for American genotype was done at >90th, >80th, >70th, >60th and <60th percentiles for better visualization. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



## Figure 2

Dispersal pattern of DENV-2 cosmopolitan genotype. The spatio-temporal dispersal pattern was determined by using Bayesian Stochastic Search Variable studies (BSSVS) analyses of 929 complete E gene sequences. The analysis demonstrates dispersal links of cosmopolitan genotype. Only the links with Bayes Factor (BF) >3 and posterior probability  $\Rightarrow 0.75$  are shown. The strength of links was categorized into very strong (BF>30) and decisive (BF>100) links based on the BF values. Each country was represented by a two letter code. AU=Australia, BD=Bangladesh, BF=Burkina Faso, BR=Borneo, CN=China, IN=India, ID=Indonesia, KE=Kenya, MM=Myanmar, MY=Malaysia, PA=Pacific Islands, PK=Pakistan, PG=Papua New Guinea, PH=Philippines, SA=Saudi Arabia, SG=Singapore, LK=Sri Lanka, TH=Thailand, TP=East Timor, TW=Taiwan, TZ=Tanzania, VN=Vietnam. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its

authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



**Figure 3**

Bayesian phylogeny of DENV-2 complete coding sequences. The maximum clade credibility (MCC) tree was constructed in BEAST 1.10.4 version, using general time reversible (GTR) substitution model, relaxed random clock model and Bayesian skyline prior. The analysis included 380 complete coding sequences of the cosmopolitan genotype and a relatively small reference dataset to represent Asian I (n=10), Asian II (n=10), Asian-American (n=8), American (n=8) and sylvatic (n=3) genotypes. The numbers on branches represent median node heights in years.

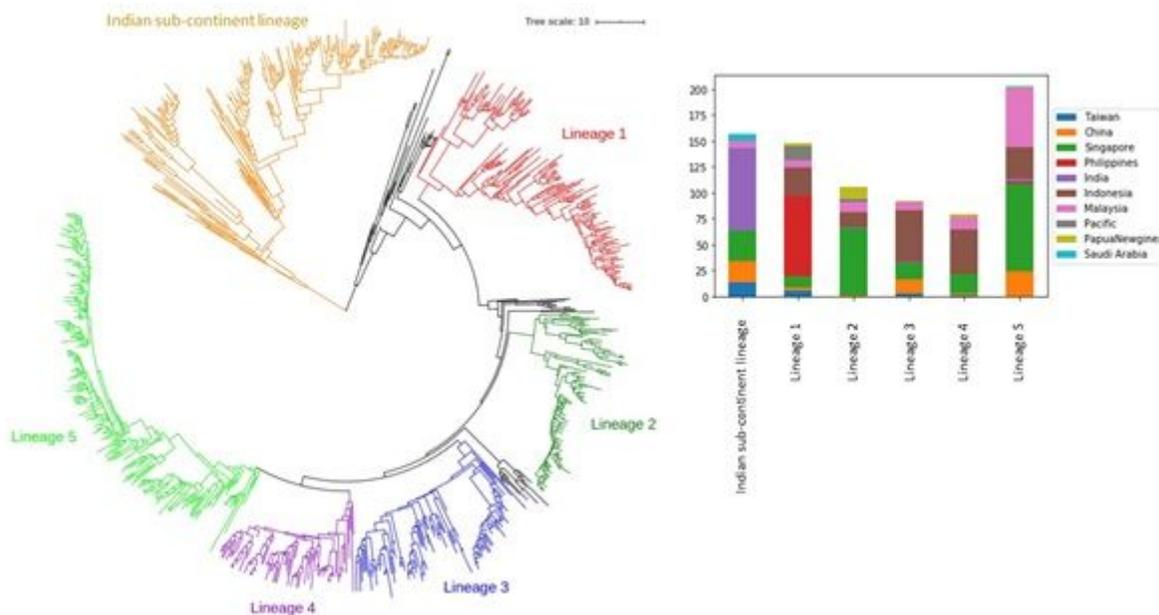


Figure 4

Phylogenetic analysis of E gene sequences of cosmopolitan genotype. Distinct lineages are shown in different colours. The inset bar chart represents geographical distribution of each lineage.

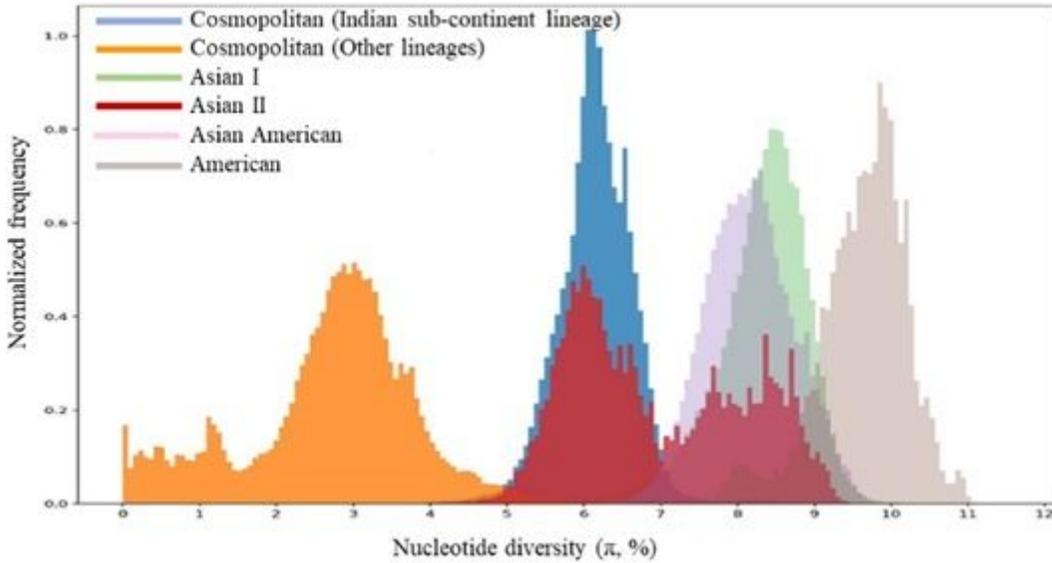


Figure 5

Pairwise nucleotide diversity ( $\pi$  index) comparison between cosmopolitan (Indian sub-continent lineage and other lineages) and other known genotypes of DENV-2. Each colour represents a different genotype of DENV-2.

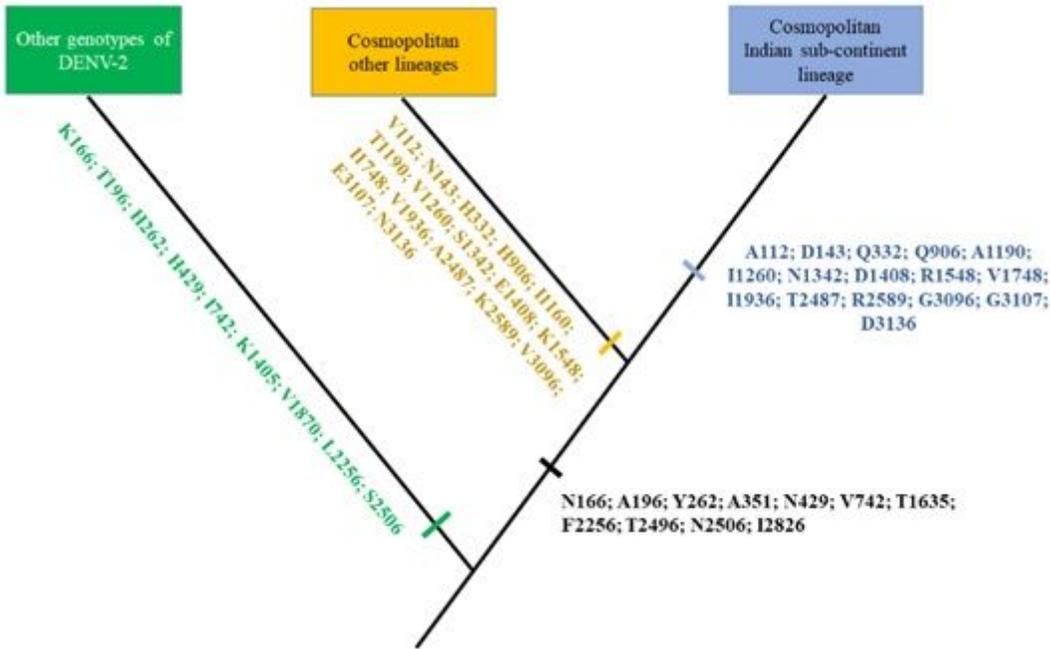


Figure 6

Schematic representation of unique amino acid signatures between Indian sub-continent and other lineages of cosmopolitan genotype. Amino acids are shown in the single letter code, with corresponding polyprotein position.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.docx](#)
- [SupplementaryTable2.docx](#)
- [SupplementaryTable3.docx](#)