International Journal of Microbiology Research
ISSN: 0975-5276 & E-ISSN: 0975-9174, Volume 13, Issue 4, 2021, pp.-1953-1962.
Available online at https://bioinfopublication.org/pages/jouarchive.php?id=BPJ0000234

Original Article

TRACKING OF GENETIC DIVERSITY, EVOLUTIONARY DYNAMICS AND ANALYSING THE DIFFERENTIAL SELECTION PRESSURE ACTING UPON DENGUE SEROTYPE 1 AND 3 IN A HYPERENDEMIC AREA OF EASTERN INDIA

SAPTAMITA GOSWAMI, RINKU CHAKRABORTI, APARNA CHOWDHURY, MEHEBUBAR RAHMAN*, SRIMA ADHIKARI, BISHAL GUPTA, SAIANTANI MONDAL AND BHASWATI BANDYOPADHYAY

Department of Topical Medicine, School of Tropical Medicine, 108, C.R. Avenue, Kolkata, 700073, West Bengal, India
*Corresponding Author: Email - rmehbub@gmail.com

Received: September 01, 2021; Revised: September 25, 2021; Accepted: September 26, 2021; Published: September 30, 2021

Abstract- Background: The burden of dengue infection in India has increased at an alarming rate as India accounts for third of global dengue infections. Kolkata is an endemic region for dengue infection. The present study describes the evolutionary analysis of dengue virus serotype 1 and 3 strains by analysing the C-prM-E region of circulating viruses in Kolkata. Methodology: C-prM-E region sequencing was performed in 7 DENV1 and 7 DENV3 strains from dengue hemorrhagic fever. MEGA software was used to develop the maximum likelihood tree. Bayesian phylogenetic analysis was done using the best fit model for each dataset. Selection pressure on structural genes was determined using the Datamonkey online platform. Result: 5 DENV1 strains grouped with American/African and 2 with Asian genotype. All DENV-3 strains were clustered with Genotype III. Mutations at the B and T cell epitopes were revealed. The nucleotide substitution rate of DENV1 was 7.42 $\times 10^{-4}$ substitutions/site/year and DENV3 was 7.19 $\times 10^{-4}$ substitutions/site/year. TMRCA of DENV1 and DENV3 viruses was estimated 132 years and 107 years respectively. Selection pressure analysis revealed that purifying negative selection was the main driving force acting on dengue virus evolution. Conclusion: The study on viral genetic diversity and evolutionary aspects will be useful for the continuous monitoring of disease burden, viral epidemiology as well as for the planning of proper prophylactic measures to control the spread of dengue infection. Several mutations at the antibody binding sites of the envelope region may help the virus to evade the host immune system.

Keywords- *DENV1, DENV3, MCC tree, Substitution rate*

Citation: Saptamita Goswami, *et al*., (2021) Tracking of Genetic Diversity, Evolutionary Dynamics and Analysing the Differential Selection Pressure Acting upon Dengue serotype 1 and 3 in a Hyperendemic Area of Eastern India. International Journal of Microbiology Research, ISSN: 0975-5276 & E-ISSN: 0975-9174, Volume 13, Issue 4, pp.-1953-1962.

Introduction

Dengue fever (DF) is one of the most prevalent arthropod-borne acute viral infections around tropical and subtropical countries. Approximately 3.6 billion people live in dengue-endemic regions worldwide and 400 million people suffer from dengue fever each year [1]. Dengue virus is enveloped single-stranded, positive-sense RNA virus. It is classified into four serotypes (DENV1-4). Each serotype shares 65-70% nucleotide sequence similarity. Each serotype is further divided into several genotypes based on 6-8% nucleotide differences [2, 3]. DENV infection causes a wide spectrum of clinical outcomes, from mild, self-limiting dengue fever to severe dengue (DS) with fatal outcomes. Circulation of several serotypes and genotypes has the potential to cause severe dengue [4-6].

Kolkata is considered as a hyperendemic region due to the circulation of multiple DENV serotypes. A large outbreak was reported in 2012 [7]. We have reported the circulation of all four serotypes during the period 2014-2017 [8]. Studies on genetic analysis based on the E region sequence from this hyperendemic region are not available to date. No study has been conducted about the evolutionary aspects of the DENV circulating in Kolkata, the Eastern part of India, in the recent past. In our present study, we will discuss the aspects of evolutionary time-scale and selection pressure analysis of circulating DENV1 and -3 in Kolkata, India by sequencing the C-prM-E region. The study based on DENV2 and DENV4 has been published elsewhere.

Materials and Methods

Sample data

Seven isolates for both DENV1 and DENV3 from separated serum samples of dengue hemorrhagic fever patients (categorized according to WHO criteria) were collected in the School of Tropical Medicine (STM), Kolkata during 2014–2017 dengue epidemics were employed in this study. This study has been approved by the Ethics Committee of the institute.

Viral RNA extraction

Viral RNA was extracted from 200µl serum samples using a QIAamp Viral RNA extraction kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The RNA was eluted in 50µl nuclease-free water and stored at -80ºC.

Reverse transcription-polymerase chain reaction (RT-PCR)

To amplify the C-prM-E region one-step RT-PCR was performed by Qiagen one-step RT-PCR kit (Qiagen, Hilden, Germany) as manufacturer's protocol. A total of six (for DENV1) and five (for DENV3) overlapping amplicons spanning the C-prM-E region were amplified using 10 primers [Table-1]. For RT-PCR, DN5URF/D1-1217R and D1-1125F/D1-2630R were used to amplify DENV1 sequences and DN5UTRF/D3-1268R and D3-1186F/ D3-2621R were used to amplify DENV3 sequences [9, 10]. The PCR products were observed in 1.5% agarose gel stained with ethidium bromide.

Sequencing Reaction

CleanSweep PCR purification kit (ThermoFisher Scientific, Waltham, MA, USA) were used to purify the PCR products. BigDye Terminator Cycle Sequencing Ready Reaction kit (v1.1) was used for the sequencing reactions

(Applied Biosystem, Foster City, CA, USA) containing ~25ng of purified PCR product and 3.2pmol of respective forward/reverse primers [Table-1]. The sequencing product was purified and loaded on 3500Dx Genetic Analyser (Applied Biosystem, Foster City, CA, USA).

Table-1 *Primers used for DNA sequencing of DENV1 and DENV3 C-prM-E region [9,10]*

Primer name	Sequence (5' to 3')	Genome region
Primers for sequencing C-PrM-E region of DENV1*		
DN5URF	AGT TGT TAG TCT ACG TGG ACC G	5'UTR (1–22)
D1-545F	ATT GCG ATG GAT TTG GGA GAG	PrM (545–565)
D1-631R	GTC AAC GTC ATC TGG TTC CG	PrM (612–631)
D1-1125F	AAA TAT CAA ACA CCA CCA CCG	E (1,125–1,145)
D1-1217R	TTC GTC GRC ACA CAA AGT TCG	E (1,197–1,217)
D1-1546F	ATC ATG GCT TGT CCA CAA AC	E (1,546–1,565)
D1-1629R	TTC CAA GTC TCT TGG GAT GT	E (1,610–1,629)
D1-2044F	CCA CCY TTT GGT GAG AGC TAC	E (2,044–2,065)
D1-2123R	TGC TTC CYT TCT TGA ACC AGC	E (2,103–2,123)
D1-2583R	CAC ACA CCC TCC TCC CAT GCC	NS1(2,563–2,583)
D1-2630R	TTT GYT TCC ACA TGA TGT TCT C	NS1(2,609–2,630)
Primers for sequencing C-PrM-E region of DENV1#		
DN5UTRF	AGT TGT TAG TCT ACG TGG ACC G	5'UTR (1–22)
D3-541F	CAA CAT GTG CAC ACT CAT AGC	PrM (529–549
D3-636R	CAG CAG TCA ATG TCT TCA GG	PrM (617–636)
D3-1186F	GGA GCA GGA CCA GAA CTA C E	(1185–1204)
D3-1268R	TCC CTT GCC AAA CAA ACC AC	E (1258–1267)
D3-1568F	TTT GAC CTA CCY CTA CCA TGG	E (1568–1588)
D3-1695R	TGC GAT CCA AGG ACT ACT ACT TCT TG	E (1669–1694)
D3-2036F	GAA CCT CCT TTT GGG GAA AG	E (2036–2055)
D3-2149R	TCT GGC AGT GGC CTC GAA C	E (2131–2149)
D3-2621R	GCT TCC ACA AGA GGT TCT CCA TTC	NS1 (2598–2621)

*Numbering from GenBank accession number AF180818
#Numbering from GenBank accession number M93130

Sequence and Phylogenetic Analysis

Nucleotide sequences were aligned and analysed by BioEdit software (v7.2.5). Phylogenetic analysis was conducted by MEGA (version 6) [11]. A maximum-likelihood phylogenetic tree for each serotype was constructed following the GTR model of nucleotide substitution with gamma-distribution rates among invariant sites (with 4 categories) with 1000 bootstrap replicates.

Bayesian MCMC Analysis

Bayesian Markov chain Monte Carlo (MCMC) approach, as implemented in the BEAST package v1.10.4 [12] was used to determine the rate of nucleotide substitution and time to the most recent common ancestor (TMRCA) of both serotypes. The best fit model for nucleotide substitution analysis was chosen by Bayesian information criterion (BIC) using jModelTest v2.1.10 [13]. The generalised time-reversible model (GTR) with complex G+I site heterogeneity model (four gamma categories) was the best fit model for both DENV1 & -3. MCMC chain was run for 30,000,000 steps and sampled at every 3000 steps. The best fit evolution model was detected by the Bayes factor [Log marginal likelihood (M1)-Log marginal likelihood (M2)>1, M1 model favoured]. Log marginal likelihoods were determined by path sampling/stepping stone sampling. Each analysis was performed in two separate runs and the resulting files were combined using LogCombiner 1.10.4 with 10% burn-ins for posterior probabilities. The resulting log files were analysed in the program Tracer v1.7.1 (ESS>200). The uncertainty in the parameter estimates was determined by a 95% HPD interval. The maximum clade credibility (MCC) tree was generated by Tree Annotator v1.10.4 and the resulting tree file was visualized in the program FigTree 1.4.4.

Recombination and Selection Pressure Analysis

Recombination analysis and selection pressure analysis of the C-prM-E region was done by using HyPhy software package under Datamonkey open-source web-server (www.datamonkey.org) incorporating the GTR model of nucleotide substitution, with a maximum-likelihood phylogenetic tree. Two datasets were used for DENV1 selection pressure analysis (American/African genotype;n=45 and Asian genotype;n=20). For DENV3 one dataset consisting of genotype III isolates (n=46) was constructed. Closely related sequences (>99.9% identity) were eliminated while constructing the datasets. The non-synonymous to synonymous substitution per site (dN/dS ratio) were estimated using four different approaches such as fast unbiased Bayesian approximation (FUBAR), mixed-effects model of evolution (MEME), single-likelihood ancestor counting (SLAC) and fixed effects likelihood (FEL). Significance levels recommended by the Datamonkey server was p<0.1 for SLAC, FEL, MEME and posterior probability>0.9 for FUBAR.

Result

Phylogenetic Analysis

DENV1 Genotype

Seven strains (MT126435-MT126441) of the DENV1 virus were sequenced (49-2373bp, n~2324bp) in this study [Table-2]. Sequences were aligned with reference DENV1 sequences to construct a Maximum-likelihood tree (numbering based on the prototype NGC 44 strain; U88535). Five isolates were clustered with American/African genotype (GIII); two isolates from 2017 were clustered with Asian genotype (GI) [Fig-1A]. Isolates clustered with American/African genotype were further subdivided into two clades CI and CII. CI isolate, STM629 was clustered with Indian isolates from Bangalore, Karnataka circulating in 2016. CII strains made a cluster among themselves. A single amino acid change *i.e.* S129T was observed between CI and CII. The pair-wise comparison of DENV1 sequences with other Indian isolates revealed 96%-99% similarity. Two strains of Asian genotype, STM1910 and STM1771, were clustered with previously reported Indian isolates from South India. These isolates displayed 97% sequence similarity with the south Indian strain.

Table-2 *Accession Number of sequenced DENV1 and DENV3 isolates C-prM-E region*

Sample ID	Year of Isolation	Accession Number	
STM629	2014	MT126435	DENV1
STM20778	2015	MT126436	
STM1322	2016	MT126437	
STM1369	2016	MT126438	
STM1771	2017	MT126439	
STM1818	2017	MT126440	
STM1910	2017	MT126441	
STM20584	2014	MT224915	DENV3
STM20616	2015	MT224916	
STM20631	2015	MT224917	
STM20974	2016	MT224918	
STM1142	2017	MT224919	
STM1531	2017	MT224920	
STM1940	2017	MT224921	

Several radical amino acid replacements were revealed in the 774aa long C-prM-E region [Table-3]. LC46M mutation was in the T-cell epitope region and IE394 was in the B-cell epitope region [14]. Six mutations (D317N, T368A, I394L, T441I, E482K, and K483E) were from the EDI region which is the interaction site for human monoclonal antibody (hMAb) 1F4 [15]. Mutations in hMAb-14c10 interacting sites in EDII and EDIII (M577V, F617I, S629T, A694T, V660I, I716V, I719V) were also found out [16].

DENV3 Genotype

All the DENV3 isolates were clustered in genotype III (MT224915- MT224921) [Table-2] [Fig-1B]. The isolates were clustered with lineage C and D circulated in India. Under lineage C, the newly sequenced isolates formed two clades CI and CII. Isolates from 2014 and 2017 clustered with lineage C and D while 2015-16 isolates (STM-20631,-20616,-20974) clustered with lineage D. Lineage C differs from lineage D at V109L. All seven DENV3 sequences showed 97%-100% similarity among them as well as with other Indian isolates. Comparison of 774 amino acid sequence with prototype strain M93130 revealed several mutations present in the study strains [Table-3]. Three mutations *i.e.* N383H, K505E, I732V at the envelope region were from the B-cell epitope region. EDI and EDII are recognized as the epitopes for DENV3 neutralizing antibody 5J7 [17]. Eleven mutations were found out in this region which may play a vital role in disease severity. Stem-loop A (SLA) at the 5'UTR region is responsible for RNA dependent RNA polymerase activity of NS5 [18]. A G-A mutation was observed within 2-69bp *i.e.* SLA site of 5'UTR region.

Bayesian MCMC Analysis

DENV1 strains

Bayesian evolutionary analysis of DENV1 strains was performed using the GTR+G+I model. Uncorrelated relaxed lognormal clock with skyline coalescent tree prior was chosen as best fit model as per the Bayes factor [Table-4]. The Maximum Clade Credibility Tree was constructed [Fig-2]. The root of the tree was shown to be 132 years old [95-168 years 95% HPD, 1884 (1849-1922)]. Time to the most recent common ancestor (TMRCA) of India II lineage of American/African genotype (genotype III) was shown to be 81 years old (62-104 years, 95% HDP). The TMRCA of newly isolated DENV1 strains (genotype III) was calculated to be 23 years (14-30years, 95% HDP). TMRCA of Asian genotype circulating was 70 years (44-95 years, 95% HDP) while the TMRCA of the Asian genotype circulating in India was 25 years (21-32years, 95% HDP). The mean nucleotide substitution rate was to be 7.42 $\times 10^{-4}$ substitutions per site per year (95% HPD, 5.89×10^{-4} to 9.29 $\times 10^{-4}$). MCC tree revealed a close association of Clade II isolates with Singapore strain isolated in 2016 (96-97.5% nucleotide similarity).

Table-3 *Amino acid substitutions in C-prM-E region of DENV1 and DENV3 isolates*

	DENV1		DENV3
Capsid	AM/AF	Asian	
	V26G	**S24P**	**R35K**
	L46M	**S39L**	M108I
	G70S		**T112A**
	N90S		L109V
prM	S129T	**A221N/T**	NIL
	T173A		
	E186D		
	K232R		
Envelop	**D317N**	**E304V**	I361V
	T368A	**E386G**	S404P
	I394L	**I425T**	H412Y
	T441I	**A426G**	S444P
	E482K	**A427D**	**A449T**
	K483E	**I443T**	**K505E**
	M577V	**N474S**	**T550N**
	F617I	**L514Q**	**K571E**
	S629T	**F515D**	**L581T**
	A694T	**A549E**	***K663N***
	V660I	**L550I**	**R671K**
	I716V	**T552M**	I732V
	I719V	**F553S**	V769A
		N564K	V438I
		I565C	**R488K**
		I600V	
		Q603R	

Note: The amino acid site involving hydrophobic to hydrophilic amino acid substitutions or vice versa are written in bold font. The amino acid site involving substitutions of neutral to charged residues or vice versa are written in bold and italics font. The amino acid positions are in respect to prototypes; DENV1-U88535(AM/AF) KJ755855 (Asian), DENV3-M93130

Table-4 Log Marginal Likelihoods analysis by Path Sampling/Stepping Stone sampling for DENV1 & DENV3

Model	DENV1	DENV3
Strict Clock, Constant Population	-15475.887	-10359.069
Strict clock, Skyline	-15512.582	-10360.873
Uncorrelated Relaxed Lognormal Clock, Constant Population	-15366.231	-10108.546
Uncorrelated Relaxed Lognormal Clock, Skyline	-15356.905	-10110.911
Uncorrelated Relaxed Exponential Clock, Constant Population	-15431.179	-10184.311
Uncorrelated Relaxed Exponential Clock, Skyline	-15416.243	-10181.218

DENV3 strains

Bayesian evolutionary analysis of DENV3 strains was performed using the GTR+G+I model. Uncorrelated relaxed lognormal clock with constant speciation model was chosen as best fit model as per the Bayes factor [Table-4]. The Maximum Clade Credibility (MCC) [Fig-3] tree supported the result of the ML tree. The mean nucleotide substitution rate was to be 7.19 $\times 10^{-4}$ substitutions per site per year (95% HPD, 5.56×10^{-4} to 8.83 $\times 10^{-4}$). The root of the tree was shown to be of 107 years old [96-120 years 95% HPD 1912 (1899-1923)]. The most recent common ancestor of genotype III was shown to be of 75 years (65-83 years, 95% HDP). TMRCA of newly sequenced DENV3 isolates was of 25 years.

Selection pressure analysis

DENV1

There was no evidence of recombination found in GARD analysis for all DENV1 data sets (AM/AF; n=45, Asian; n=20) [Table-5]. DENV1 codons (n=774) were strongly under negative selection [Table-6]. Within AF/AF genotype six sites and within Asian genotype seven sites were under episodic positive selection, identified by MEME method [Table-7]. Non-synonymous mutation sites at the envelope region of DENV1 Asian genotype E515, E552 and E565 were under positive selection.

Table-5 *Reference sequences used for DENV1 selection pressure analysis*

Accession Number	Country Name	Year of Isolation	Genotype
JN903579	India	2008	AM/AF
JN903578	India	2007	AM/AF
JQ922548	India	2005	AM/AF
JQ922545	India	1982	AM/AF
JQ917404	India	2009	AM/AF
JN903578	India	2007	AM/AF
JN903581	India	2009	AM/AF
JQ922544	India	1963	AM/AF
KJ189303	Colombia	1998	AM/AF
GQ868562	Colombia	2005	AM/AF
JN819410	Venezuela	2005	AM/AF
KF973461	Nicaragua	2012	AM/AF
KJ189348	Mexico	2011	AM/AF
KJ189312	Mexico	2008	AM/AF
FJ850070	Brazil	2000	AM/AF
KX618706	India	2014	AM/AF
KF289073	India	1956	AM/AF
MF033253	Singapore	2016	AM/AF
KU509255	India	2011	AM/AF
JQ692085	India	2010	AM/AF
MN923082	China	2019	AM/AF
MN933653	China	2013	AM/AF
JQ917404	India	2009	AM/AF
KF289072	India	2011	AM/AF
JQ692085	India	2010	AM/AF
MK858119	India	2016	AM/AF
MK858139	India	2016	AM/AF
MF033204	Singapore	2012	AM/AF
MK588396	India	2016	AM/AF
MG721060	India	2016	AM/AF
AY732474	Thailand	1980	AM/AF
DQ285562	Comoros	1993	AM/AF
KX618705	India	2014	AM/AF
KP406801	South Korea	2004	AM/AF
MK791752	India	2016	AM/AF
KF289073	India	1956	AM/AF
JQ922546	India	1971	AM/AF
MK858128	India	2016	AM/AF
KU509255	India	2011	AM/AF
JF815192	India	2010	AM/AF
KP256367	India	2013	AM/AF
KU365900	Taiwan	2014	ASIAN
AB189120	Indonesia	1998	ASIAN
MN018336	Malaysia	2016	ASIAN
AY726555	Myanmar	1998	ASIAN
FJ469907	Singapore	2003	ASIAN
GQ868602	Philippines	2004	ASIAN
JN054255	Srilanka	2010	ASIAN
JN638336	Thailand	1986	ASIAN
GU131792	Vietnam	2008	ASIAN
AB074761	Japan	2001	ASIAN
KJ755855	India	2013	ASIAN
EU081258	Singapore	2005	ASIAN
AY722803	Myanmar	1998	ASIAN
AY722802	Myanmar	1996	ASIAN
HQ891316	Srilanka	2009	ASIAN
EF025110	China	2002	ASIAN
GU131923	IPC	2005	ASIAN

Fig-1A Maximum likelihood (ML) phylogenetic tree of DENV strains. The tree was constructed using MEGA (v6) software with bootstrap support of 1000 replicates. The Each strain is denoted by Genbank accession number followed by country of origin and the year of isolation. Numbers on nodes indicate bootstrap support generated by 1000 replicates. Bootstrap values of >70 is shown. (1A) DENV1 strains sequenced in the study are marked by black circle, (1B) DENV3 strains sequenced in the study are marked by black diamond.

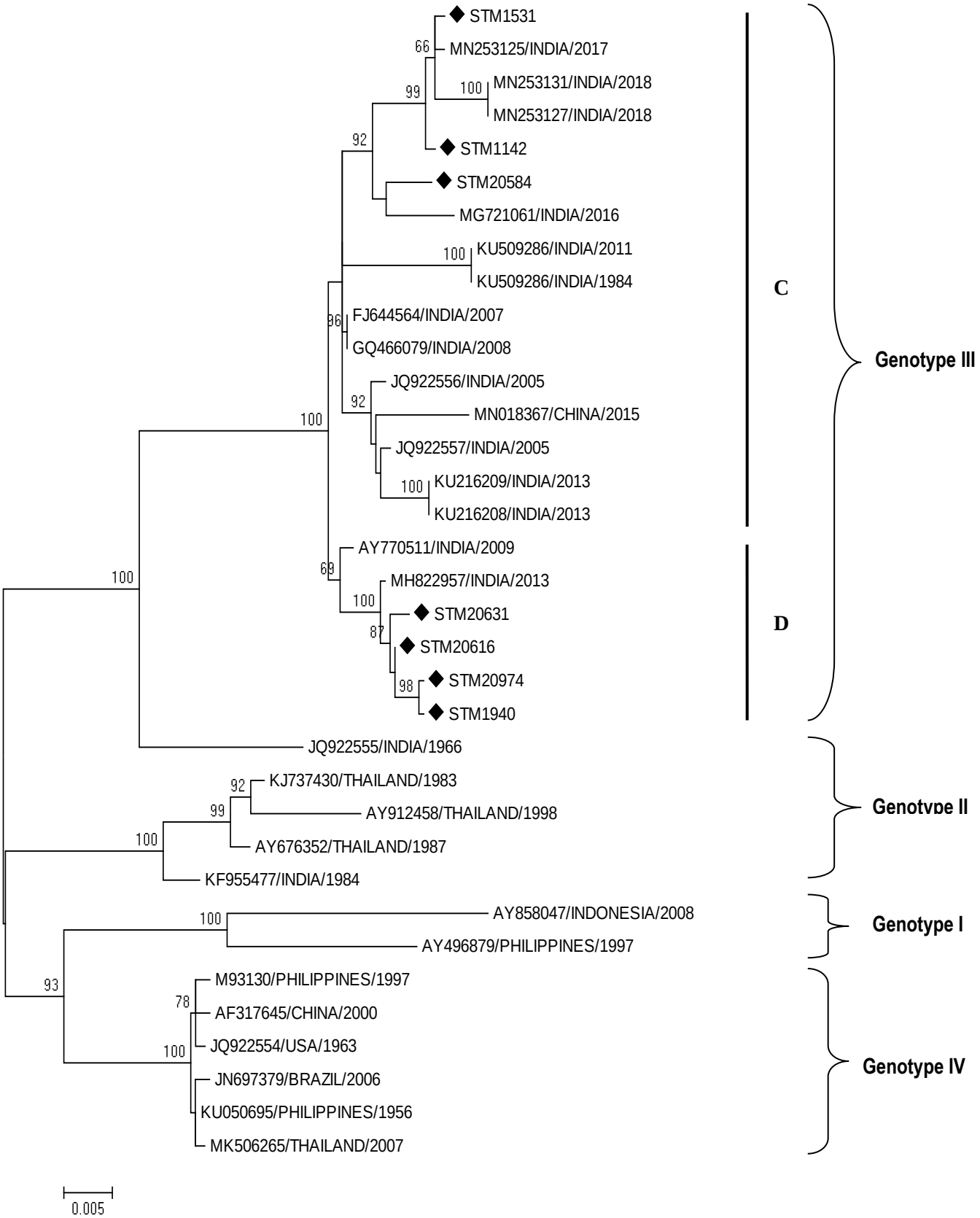

Fig-1B Maximum likelihood (ML) phylogenetic tree of DENV strains. The tree was constructed using MEGA (v6) software with bootstrap support of 1000 replicates. The Each strain is denoted by Genbank accession number followed by country of origin and the year of isolation. Numbers on nodes indicate bootstrap support generated by 1000 replicates. Bootstrap values of >70 is shown. (1A) DENV1 strains sequenced in the study are marked by black circle, (1B) DENV3 strains sequenced in the study are marked by black diamond.

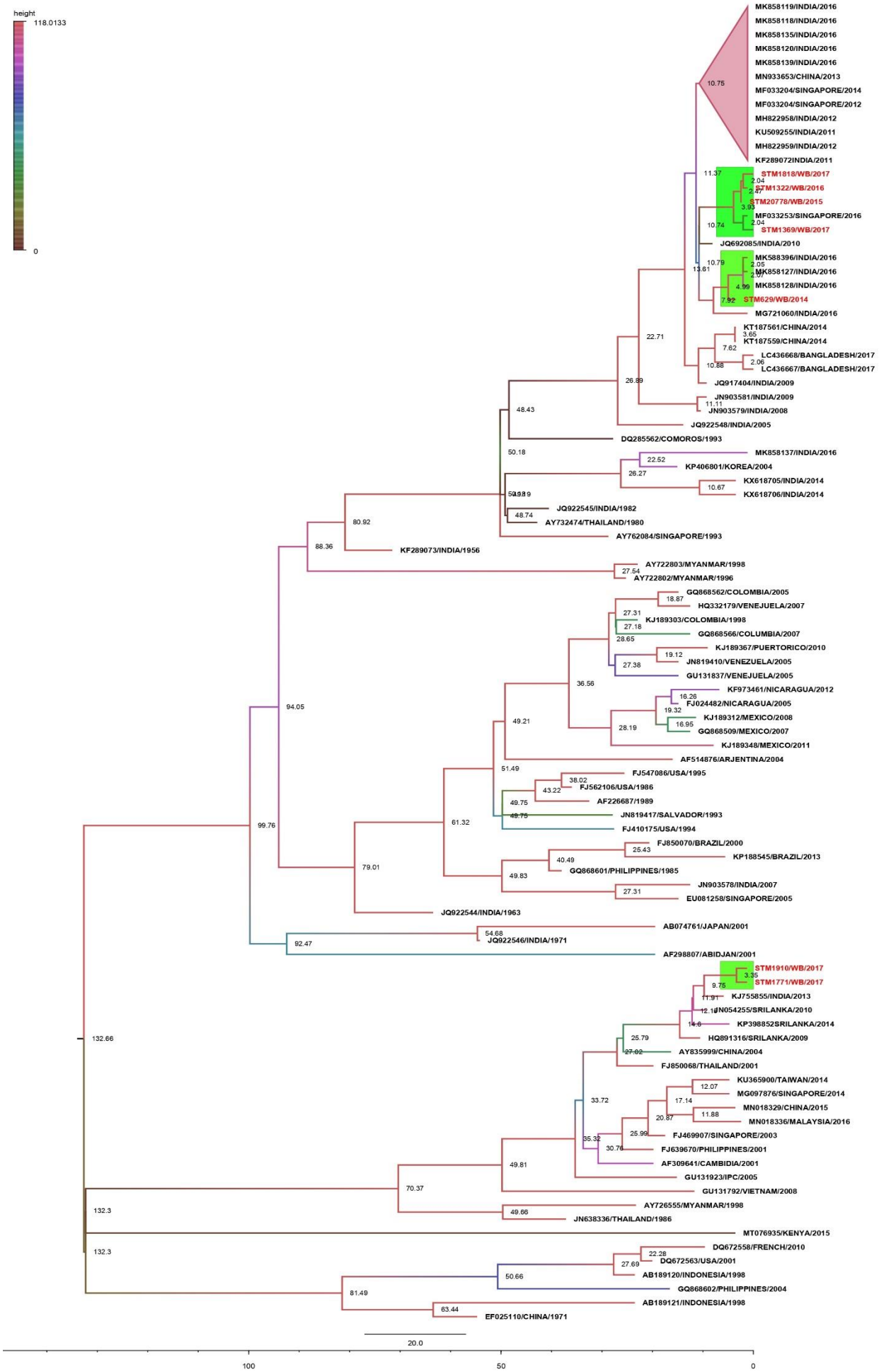

Fig-2 Bayesian Maximum Clade Credibility tree (MCC) of DENV1 isolates. Tree derived with the best fit model. Node ages are shown at each node.

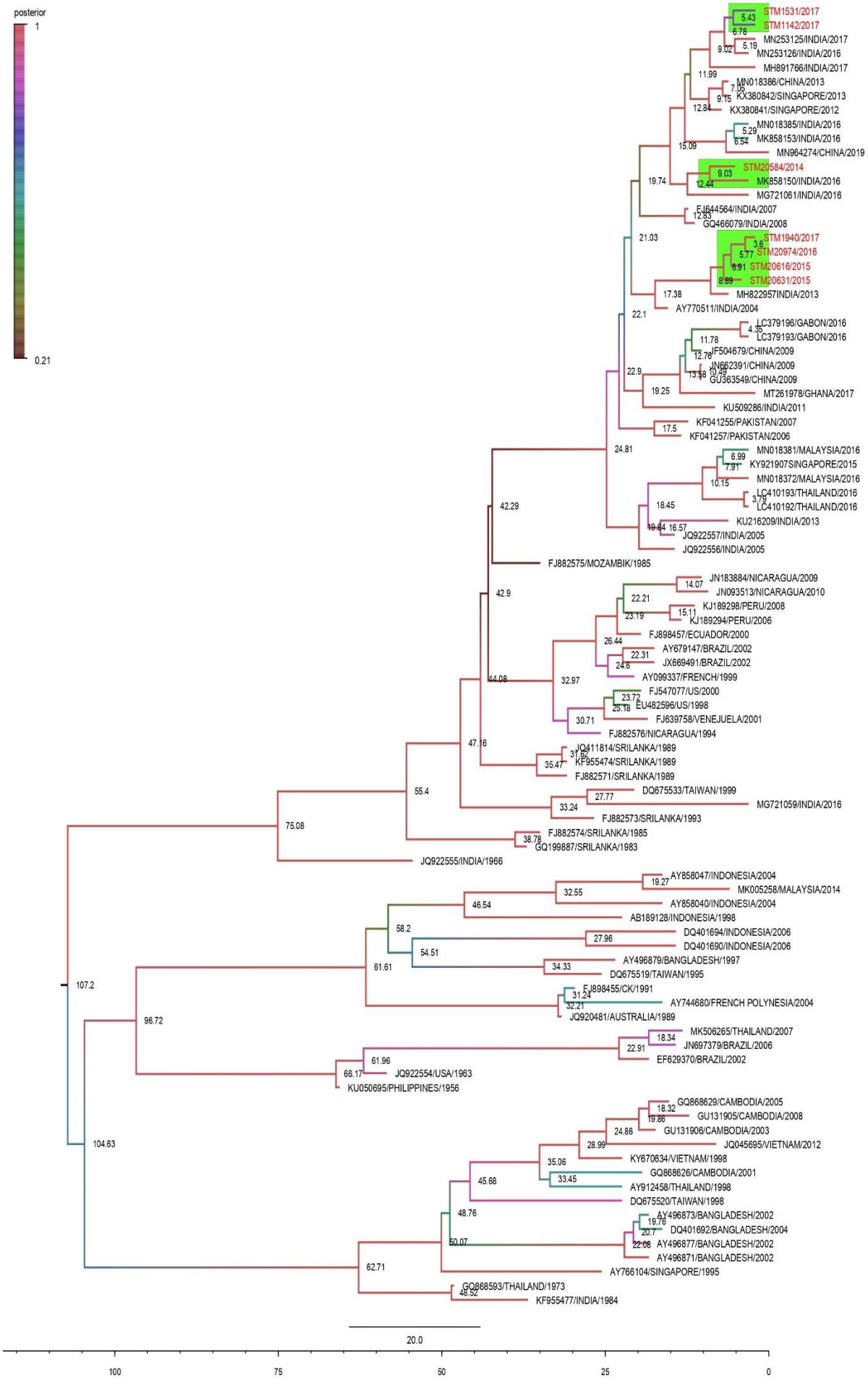

Fig-3 Bayesian Maximum Clade Credibility tree (MCC) of DENV3 isolates. Tree derived with the best fit model. Node ages are shown at each node.

Table-6 Negatively selected sites of DENV1 and DENV3 (C-prM-E) by SLAC, REL, FEL and FUBAR methods

Serotype	Genotype	SLAC	FEL	FUBUR
DENV1	AM/AF	67	173	288
	Asian	69	216	359
DENV3	Genotype III	178	291	450

Table-7 *Parameters from selection pressure analysis of DNEV1 and DENV3 (C-prM-E) gene using the SLAC, REL, FEL and FUBAR methods*

Serotype	Codon	SLAC		FEL		MEME		FUBAR	
	(protein)	dN/dS	p-value	dN/dS	p-value	dN/dS	p-value	Bayes factor	Posterior Probability
AM/AF (n=45)	C54	Infinity	0.36	Infinity	0.22	**4.30**	**0.05**	1.74	0.312
	C141	1.00	0.74	0.97	0.97	**7.05**	**0.01**	1.10	0.73
	M274	3.36	0.31	2.28	0.49	**6.96**	**0.01**	1.83	0.62
	M278	1.36	0.65	0.46	0.61	**18.02**	**0.03**	0.14	0.03
	E312	0.75	0.79	0.70	0.70	**6.79**	**0.02**	0.87	0.77
	E437	0.60	0.82	0.47	0.54	**8.29**	**0.01**	0.95	0.76
ASIAN (n=20)	C95	Infinity	0.97	Infinity	0.83	**6.85**	**0.01**	3.073	0.444
	C96	0.81	0.79	0.88	0.92	**5.25**	**0.03**	1.765	0.314
	M112	Infinity	0.29	Infinity	0.10	**3.27**	**0.09**	16.695	0.812
	E406	Infinity	0.74	Infinity	0.42	**3.95**	**0.06**	3.731	0.492
	E515	0.27	0.97	0.30	0.21	**5.06**	**0.04**	0.066	0.017
	E552	2.27	0.54	23.87	0.60	**6.16**	**0.02**	4.365	0.531
	E565	0.42	0.34	0.41	0.41	**6.43**	**0.02**	0.336	0.080
DENV3 (n=46)	M115	Infinity	1.00	0.574	0.586	**9.43**	**0.00**	0.97	0.24
	E332	Infinity	1.00	0.316	0.419	**6.49**	**0.02**	1.07	0.26
	E336	0.526	0.634	0.484	0.509	**10.39**	**0.00**	0.34	0.10
	E339	1.06	0.400	0.279	0.365	**4.37**	**0.05**	0.06	0.02
	E340	1.11	0.308	0.598	0.568	**7.79**	**0.01**	0.18	0.05
	E357	1.25	0.383	0.329	0.425	**12.67**	**0.00**	0.16	0.05
	E407	Infinity	1.00	0.904	0.934	**10.88**	**0.00**	1.02	0.25
	E440	Infinity	1.00	0.847	0.934	**8.02**	**0.01**	1.11	0.27
	E450	Infinity	1.00	1.566	0.594	**8.73**	**0.01**	0.98	0.24
	E491	0.207	0.208	Infinity	0.923	**8.64**	**0.01**	0.18	0.05
	E497	Infinity	1.00	0.361	0.182	**3.77**	**0.07**	0.86	0.72
	E550	Infinity	1.00	Infinity	0.196	**6.90**	**0.01**	3.55	0.40
	E551	Infinity	1.00	**Infinity**	**0.010**	**9.22**	**0.00**	0.90	0.233
	E552	Infinity	1.00	1.236	0.849	**17.72**	**0.00**	1.06	0.25
	E553	Infinity	1.00	Infinity	0.602	**4.92**	**0.04**	1.14	0.27
	E555	1.11	0.304	0.195	0.231	**4.49**	**0.05**	1.18	0.05
	E578	Infinity	1.00	Infinity	0.160	**8.54**	**0.01**	0.91	0.71
	E584	1.17	0.285	**Infinity**	**0.076**	**3.15**	**0.10**	0.18	0.05
	E636	Infinity	1.00	1.068	0.983	**5.97**	**0.02**	0.95	0.71
	E686	Infinity	1.00	0.782	0.842	**3.50**	**0.08**	1.24	0.65
	E717	0.618	0.611	Infinity	0.586	**5.46**	**0.03**	0.66	0.78

Sites found under positive selection by at least one method were shown in bold. C-Capsid, M-Membrane, E-Envelope.
P value <0.1 and Bayes factor >0.9 indicates statistically significant evidence for positive selection.

DENV3

There was no evidence of recombination found in the GARD analysis of the dataset (n=46) [Table-8]. MEME revealed the presence of 21 episodic positive selection sites while in FEL analysis two positively selected sites were reported [Table-7]. E551 and E584 are from E/DI and E/DII region which is the epitope for neutralizing human monoclonal antibody (hMAb).

Discussion

Kolkata, the economic centre of eastern India, is a hyperendemic region for dengue infection. The battle against dengue infection is a never-ending phenomenon since 1964. In every 2-3 years incidents of dengue infection surge to a new level due to circulation of all four DENV serotypes. In the 2010 outbreak, the dominant serotype was DENV2 [19]. Dominance shifts from DENV2 to DENV1, -2 were reported in the 2012 outbreak [7]. Further study shows the change in predominance and circulation of multiple serotypes during 2014-2017 [8]. In this study, we have discussed the molecular and evolutionary aspects of DENV1 and DENV3 serotypes circulating in Kolkata.

In this study, five newly sequenced DENV1 strains were clustered with American/African (AM/AF) genotypes. A previous study on DENV1 Indian strains reported the circulation of four lineages of AM/AF genotype [20]. The newly sequenced strains from Kolkata have been clustered with India II lineage. The presence of India II lineage is reported since 1962. The TMRCA of this lineage is of 58 years. Surprisingly two DENV1 strains have been clustered with the Asian genotype. Circulation of Asian genotype was reported in 1997-98 from North India [21]. In 2012, a large outbreak from South India was reported due to displacement of AM/AF genotype with Asian genotype [22]. The study strains isolated in 2017 show 99% similarity with the South Indian 2012 strains. Circulation of two different genotypes in the same year is a rare phenomenon as compared to the circulation of multiple serotypes which is a common event. The presence of several non-synonymous mutations throughout the C-prM-E region was evident from this study. K232R is an important non-conservative substitution at the B-cell epitope site of the membrane region. The impact of these genetic changes needs further research.

DENV3 isolates from this study were grouped with two lineages *i.e.* Lineage C and D of genotype III. Circulation of both of these lineages has been reported since 2005 [23]. A large number of non-synonymous mutations were present at the DI/II region of the envelope protein. This region is the binding epitope of human monoclonal antibody (hMAb) 5J7 [24]. RNA viruses have higher magnitude of substitutions rate than those of DNA viruses mainly due to their polymerases [25]. Another important factor contribute to high substitution rate is positive selection pressure [26]. High mutation rates allow the viruses to escape from immune surveillance of hosts. ssRNA viruses such as HIV, influenza virus, DENV are the fastest evolving viruses with rates ranging between 10^{-2} and 10^{-5} substitutions per site per year (s/s/y) [27]. The nucleotide substitution rate of DENV1 viruses was determined in the study which is similar to the rate previously reported (6.5×10^{-4} substitutions/site/ year) [20].

Table-8 *Reference sequences used for selection pressure analysis of DENV3 (genotype III)*

Accession Number	Country Name	Year of Isolation
MN253125	India	2017
MN253126	India	2016
MH891766	India	2017
MN018386	China	2013
KX380842	Singapore	2013
KX380841	Singapore	2012
MN018385	India	2016
MK858153	India	2016
MN964274	China	2019
MK858150	India	2016
MG721061	India	2016
FJ644564	India	2007
GQ466079	India	2008
MH822957	India	2013
AY770511	India	2004
LC379196	Gabon	2016
JF504679	China	2009
JN662391	China	2009
GU363549	China	2009
MT261978	Ghana	2017
KU509286	India	2011
KF041255	Pakistan	2007
MN018381	Malaysia	2016
KY921907	Singapore	2015
LC410192	Thailand	2016
KU216209	India	2013
JQ9922557	India	2005
JQ922556	India	2005
FJ882575	Mozambik	1985
JN183884	Nicaragua	2009
JN093513	Nicaragua	2010
KJ189298	Peru	2008
FJ898457	Ecuador	2000
AY679147	Brazil	2002
JX669491	Brazil	2002
AY099337	French	1999
FJ547077	US	2000
EU482596	US	1998
FJ882576	Nicaragua	1994
JQ411814	Sri Lanka	1989
FJ882571	Sri Lanka	1989
DQ675533	Taiwan	1999
MG721059	India	2016
KF955474	Sri Lanka	1989
GQ199887	Sri Lanka	1983
JQ922555	India	1966

The mean nucleotide substitution rate of DENV3 viruses determined by using a relaxed lognormal clock is comparable to the previously reported rate of 8.9×10^{-4} substitutions/site/ year [28]. Selection pressure analysis of both DENV1 and DENV3 revealed that a strong purifying selection is acting on both serotypes. This is due to evolutionary constraints as the dengue virus life cycle involves two taxonomically different hosts [3]. Envelope protein is a major determinant of viral entry, cellular tropism and the target of both humoral and cellular immune selection [29]. E312 and E437 sites in the American/African genotype and E406, E515, E552 and E565 sites in Asian genotype showed evidence of positive selection. Two sites at the E region, E551 and E584 of DENV3 were shown positively selected by two methods adopted for the analysis. These sites were not been reported previously. E312, E406, E437 were located in domain I and E515 and E552 were in the domain II region of the envelope protein. Domain II acts at low pH and involves exposing the fusion peptide on the virion surface at the time of virus entry into host cells [30]. This residue may be responsible for altering virus-host cell interactions.

Conclusion

We have reported the circulation of two different genotypes *i.e.* AM/AF and Asian genotypes of DENV1 and different lineages *i.e.* lineage C and D of DENV3 genotype III.

This study also has estimated the substitution rate of circulating strains and the number of mutations present throughout the study region. A significant drawback of this study is that only structural regions of the virus have been studied.

Application of research: This study has several findings that can help to understand the nature and evolutionary pattern of the circulating dengue virus in Eastern India.

Research Category: Dengue infection

Acknowledgement / Funding: Authors are thankful to The Department of Biotechnology (DBT), Government of India (grant number BT/PR8597/MED/29/764/2013). The authors are thankful to Dr Arup Banerjee from Translational Health Science and Technology Institute (THSTI), Faridabad, India for the valuable pieces of advice for the smooth continuation of this study. The authors are thankful to Department of Topical Medicine, School of Tropical Medicine, 108, C.R. Avenue, Kolkata, 700073, West Bengal, India

****Principal Investigator or Chairperson of research: Prof Bhaswati Bandyopadhyay**
Institute: Virology unit, School of Tropical Medicine, Kolkata, 700073, India
Research project name or number: Institute Research Project

Author Contributions: All authors equally contributed

Author statement: All authors read, reviewed, agreed and approved the final manuscript. Note-All authors agreed that- Written informed consent was obtained from all participants prior to publish / enrolment

Study area / Sample Collection: Kolkata, 700073, West Bengal

Strain name: Seven strains (MT126435-MT126441) of the DENV-1 virus and seven strains (MT224915- MT224921) of DENV3 virus

Conflict of Interest: None declared

Ethical approval: The study protocol was ethically approved by the Clinical Research Ethics Committee (CREC-STM) of School of Tropical Medicine, Kolkata. All patients (>12 years of age) enrolled in this study were provided with a written consent form. The experimental methods proceeded in compliance with approved guidelines.
Ethical Committee Approval Number: CREC-STM/195 dated 24.04.2014

References

[1] Bhatt S., Gething P.W., Brady O.J., *et al*. (2013) *Nature*, 496(7446), 504-507.
[2] Holmes E.C., Twiddy S.S. (2003) *Infect Genet Evol*., 3(1), 19-28.
[3] Weaver S.C., Vasilakis N. (2009) *Infect Genet Evol.*, 9(4), 523–540.
[4] Halstead S.B. (2007) *Lancet*, 370(9599), 1644–1652.
[5] Banerjee A., Shukla S., Pandey A.D., et at. (2017) *Transl Res*., 186, 62-78.e9.
[6] Pandey A.D., Goswami S., Shukla S., *et al*. (2017) *J Infect*., 75, 541-554.
[7] Bandyopadhyay B., Bhattacharyya I., Adhikary S., *et al*. (2013) *Hindawi Publishing Corporation*, 207580, 5.
[8] Goswami S., Rahman M., Chakraborti R., Karmakar J., Bhattacharya N., Bhaswati B. (2021) *Inter J Contem Med Res*., 8(1), A1-A9.
[9] Huang J.H., Liao T.L., Chang S.F., et al. (2007) *Am J Trop Med Hyg*., 77, 903–909.
[10] Shu P.Y., Su C.L., Liao T.L., et al. (2009) *Am J Trop Med Hyg*., 80(6), 1039–1046.
[11] Tamura K., Peterson D., Peterson N., et al. (2011) *Mol Bio. Evol*., 28, 2731–2739.

[12] Drummond A.J., Suchard M.A., Xie D., Rambaut A. (2012) *Mol Biol Evol*., 29(8), 1969–1973.
[13] Darriba D., Taboada G.L., Doallo R., Posada D. (2012) *Nature Methods*, 9(8), 772.
[14] Vaughan K., Greenbaum J., Blythe M., Peters B., Sette A. (2010) *Viral Immunol*., 23(3), 259–284.
[15] Fibriansah G., Tan J.L., Smith S.A., *et al*. (2014) *EMBO Mol Med*., 6(3), 358-371.
[16] Teoh E.P., Kukkaro P., Teo E.W., *et al*. (2012) *Sci Transl Med*., 4(139), 139ra83.
[17] de Alwis R., Smith S.A., Olivarez N.P., *et al*. (2012) *Proc Natl Acad Sci.,* 109(19), 7439-7444.
[18] Gebhard L.G., Filomatori C.V., Gamarnik A.V. (2011) *Viruses*, 3(9), 1739–1756.
[19] Sarkar A., Taraphdar D., Chatterjee S. (2012) *J Trop Med.,* Article ID 960329, 5 pages.
[20] Yergolkar P.N., Cherian S.S., Jadhav S., Raut C.G., Mourya D.T. (2017) Indian J Med Res., 146(5), 662-665.
[21] Kukreti H., Dash P.K., Parida M., Chaudhary A., Saxena P., Rautela R.S., *et al*. (2009) *Virol J.,* 6, 1.
[22] Cecilia D., Patil J.A., Kakade M.B., *et al*. (2017) *Virology*, 510, 40-45.
[23] Afreen N., Naqvi I.H., Broor S., Ahmed A., Parveen S. (2015) *PLoS one*, 10(11), e0141628.
[24] Fibriansah G., Tan J.L., Smith S.A., *et al*. (2015) *Nat Commun*., 6, 6341.
[25] Holmes Ed. (2010) *Proceedings of the National Academy of Sciences*, 107, 1742-1746.
[26] Duffy S., Shackelton L.A., Holmes E.C. (2008) *Nat Rev Genet*., 9(4), 267-276.
[27] Domingo-Calap P., Schubert B., Joly M., *et al*. (2018) *PLoS Pathog*., 14(10), e1007368.
[28] Araújo J.M.G., Nogueira R.M.R., Schatzmayr H.G., Zanotto P.M.D., Bello G. (2009) *Infect Genet Evol*., 9(4), 716–725.
[29] Modis Y., Ogata S., Clements D., Harrison S.C. (2003) *Proc Natl Acad Sci*., 100(12), 6986-6991.
[30] Guardia C.L., Lleonart R. (2014) *BioMed Res Int*, Article ID 825039.