Complete genome sequencing and evolutionary analysis of dengue virus serotype 1 isolates from an outbreak in Kerala, South India

**M. Anoop · Ashish J. Mathew · B. Jayakumar ·
Aneesh Issac · Sajith Nair · Rachy Abraham ·
M. G. Anupriya · E. Sreekumar**

Abstract In this study, dengue virus (DENV) isolates from a localized, small-scale, non-seasonal dengue outbreak were genetically characterized. The outbreak occurred during the pre-monsoon months (April–May) in a medical college campus in Kerala, South India in 2009 affecting 76 people. Analysis of 39 viral RNA positive serum samples by a serotype specific reverse-transcription polymerase chain reaction identified dengue virus serotype 1 (DENV1) as the causative strain. Formation of a distinct genetic clade was revealed in the initial phylogenetic analysis using nucleotide sequences of a partial (303 bp) Capsid-Pre-membrane protein (C-PrM) coding region of 37 outbreak strains. The sequences of these strains clustered with that of the Genotype III DENV-1 strains from India, and 32 among them formed a single major sub-clade. Whole-genome sequencing (10,693 bp) of two strains (RGCB585/2009 and RGCB592/2009) selected from this major sub-clade, and subsequent phylogenetic analysis using the full-length coding region sequence showed that the sequences grouped with that of the isolates from Thailand (1980), Comoros (1993), Singapore (1993), and Brunei (2005) among the Indo-Pacific isolates. The sequences of the two strains had a nucleotide identity of 97–98 % and an amino acid identity of 98–99 % with these closely related strains. Maximum amino acid similarity was shown with the Singapore 8114/93 isolate (99.6 %). Four mutations—L46M in the capsid, D278N in the NS1, L123I, and L879S in the NS5 protein coding regions—were seen as signature substitutions uniformly in RGCB585/2009 and RGCB592/2009; in another isolate from Kerala (RGCB419/2008) and in the Brunei isolate (DS06-210505). These four isolates also had in common a 21-nucleotide deletion in the hypervariable region of the 3′-non-translated region. This first report on the complete genome characterization of DENV-1 isolates from India reveals a dengue outbreak caused by a genetically different viral strain. The results point to the possibility of exotic introduction of these circulating viral strains in the region.

Keywords Dengue virus serotype 1 · Whole-genome sequence · Independent evolution · Maximum-likelihood analysis

M. Anoop · A. Issac · S. Nair · R. Abraham ·
M. G. Anupriya · E. Sreekumar (✉)
Rajiv Gandhi Centre for Biotechnology (RGCB), Thycaud P.O.,
Thiruvananthapuram 695014, Kerala, India
e-mail: esreekumar@rgcb.res.in

A. J. Mathew · B. Jayakumar
Department of Internal Medicine, Medical College Hospital,
Thiruvananthapuram 695011, Kerala, India

Introduction

Dengue virus (DENV) is a positive sense single-stranded RNA virus of the *Flaviviridae* family. It is the etiological agent of dengue, a major mosquito-borne febrile infection of the tropics [1]. The disease manifests either as mild Dengue fever or as severe, and sometimes life threatening, Dengue hemorraghic fever (DHF) or Dengue shock syndrome (DSS). The viral genome is approximately 11,000 basepairs (bp) long and consists of a small 5′-non-translated region (5′-NTR) of ∼100 bp; a 10,179 bp open reading frame that codes for the three structural and seven non-structural proteins; and a 3′-NTR of ∼400 bp [2]. The

structural proteins are the capsid (C), membrane (M), and envelope (E) proteins and the non-structural proteins are the NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5 proteins. The 5′-NTR of the genome is 7-methyl-guanosine capped and the 3′-NTR has no polyadenylation [2]. In nature, the virus circulates as four serotypes (DENV-1–4) and shares about 65–70 % sequence homology with each other [2]. These viruses have the highest mutation rate among the Flavivirus group [3, 4].This has led to the formation of different genotypes and lineages within each serotype [5]. Accordingly, DENV-1 comprises of five genotypes (I–V); DENV-2 has six genotypes (Southeast Asian/American, Asian I, Asian II, Cosmopolitan, American, and Sylvatic); DENV-3 has four genotypes (I–IV); and DENV-4 also consists of four genotypes (I, II, III, Sylvatic) [6]. Genetic changes observed in dengue viruses can have major impact on disease burden. A shift in circulating genotype can affect the disease severity as exemplified by the replacement of less virulent American DENV-2 genotype with more virulent Southeast Asian DENV-2 [7], and also by increased occurrence of DHF in Sri Lanka caused by the replacement of DENV-3 subtype III group A viruses with group B [8]. Some genotypes show increased viraemia during infection [9] leading to enhanced transmissibility and outbreak potential [4]. Progressive displacement of less virulent genotypes with more virulent ones during large-scale dengue infections has been reported [10]. Knowledge on the circulating virus genotypes in the dengue endemic regions has become important in disease surveillance, epidemiology, and vaccine development [11].

Conventionally, selected gene segments are used for the molecular characterization of the dengue virus strains. This approach is getting replaced by whole-genome analysis, which would help to understand the dengue disease dynamics better [12]. Among the four dengue virus serotypes, DENV-1 is a predominant strain that circulates in all major dengue prevalent countries [13]. Complete genome analysis of the DENV-1 isolates from several countries such as Paraguay and Argentina [14], Singapore [12], Brunei [15], Thailand [16], and Cambodia [17] has been published. Dengue is endemic to India [18, 19] and the previous genetic studies of dengue strains from the country were based on analysis of smaller genomic regions [20–24]. Recently, whole-genome sequence analysis of DENV-3 strains from the country has been published [25]. However, complete genome sequence analysis of other DENV serotypes has not been reported and there is a definite lack of sufficient genetic information on these strains from India.

The DENV-1 was detected in India even as early as in 1956 from Vellore, South India [26]. Subsequently, a number of outbreaks, such as the ones in 1968 [26], 1997 [27], 2006 [20], and 2008 [21], were attributed to DENV-1

infection wherein a progressive change in the disease profile from mild to severe form was observed [20]. Genetic variations in the circulating DENV-1 strains were observed during these years with emergence of a new lineage [20] and co-circulation of multiple genetic lineages of DENV-1 for several years [21]. Subsequent to the first reports on DENV-1 from South India [26], several disease outbreaks have been reported from the region [28, 29]. Confirmed cases of dengue in Kerala, located in the Southwest-coast of the peninsula, were reported for the first time in 1997 wherein presence of DENV-1, DENV-2, and DENV-4 were identified by serological analysis [28]. The most severe dengue outbreak in Kerala occurred in 2003 with 3,546 confirmed cases and 68 cases of mortality. DENV-3 was attributed as a predominant serotype along with the presence of DENV-2 [28]. Though the number of cases fell to 686 in the following year, since then there has been a steady increase in dengue cases in the state with 2,597 cases reported in 2010 that accounted for 17 deaths [19]. Molecular studies on viral strains from Kerala are lacking and in our previous study based on reverse-transcription polymerase chain reaction (RT-PCR) and nucleotide sequencing, co-circulation of multiple dengue virus serotypes and combined infections involving Genotype III of DENV-1, Genotype IV of DENV-2, and Genotype III of DENV-3 were documented [30]. Major outbreaks involving DENV-1 strains have not been reported in the region so far. With direct sea and air-links to the dengue endemic Southeast Asian countries on the Pacific Rim, and presence of thick tropical vegetation providing ideal breeding grounds for the vector Aedes mosquitoes, the region can act as a major point of entry and spread of new dengue virus strains in the subcontinent. This study describes the characterization of DENV-1 strains isolated from a focal, small-scale dengue outbreak in the state by complete genome sequencing and comparative sequence analysis.

Materials and methods

Clinical samples

A small-scale fever outbreak that occurred in a Medical college campus in Thiruvananthapuram, the capital city of Kerala state during April–May 2009 served as the source of the samples. Blood samples (2–5 ml) were collected under informed consent from patients who were admitted with fever, head-ache, retro-orbital pain, myalgia, and backache, and clinically diagnosed as Dengue in the hospital. Samples were transported to the laboratory in cold-chain and used for anti-dengue IgM and IgG detection, RT-PCR and for virus isolation.

Laboratory diagnosis by serology and RT-PCR

Presence of anti-dengue IgM and IgG antibodies in patient serum was detected using a commercial ELISA kit (IVD Research Inc, Carlsbad, USA) as per the protocol supplied along with the kit. An in-house single tube RT-PCR that can amplify RNA corresponding to a 654-bp C-PrM coding region from all the four serotypes has been described previously [30]. This was used for Dengue viral RNA detection in the patient samples. In brief, viral RNA was extracted from 140 µl of the serum samples using QIAmp Viral RNA Mini Kit (Qiagen, Germany) following manufacturer's instructions and RNA was finally eluted in 60 µl of the elution buffer. 2 µl of the RNA was used in a 25 µl single-step RT-PCR reaction containing 1× RT-PCR master mix (USB, Cleveland, Ohio) and 10 pmol each of the D1F and DencomR2 as forward and reverse primers (Table 1), and was subjected to reverse transcription at 42 °C for 30 min. This was followed by PCR with an initial denaturation at 95 °C for 5 min; and 35 cycles of amplification each with a denaturation at 94 °C for 30 s, annealing at 55 °C for 1 min, and extension at 68 °C for 1 min; and a final extension step of 68 °C for 2 min. The 654-bp amplified product obtained in the positive samples were purified by GFX gel band elution kit (GE, Amersham, USA) and were further subjected to a multiplex semi-nested PCR using D1F as the forward primer and nTS1, nTS2, nTS3, and nDen4 as the reverse primers (Table 1) for serotype identification as previously described [31, 32]. This multiplex PCR using slightly modified type-specific primers (Table 1) was designed to amplify a 489-bp fragment for DENV-1, 123 bp for DENV-2, 296 bp for DENV-3, and 395 bp for DENV-4. 25 µl PCR reaction mixture contained 1× GoTaq PCR Master mix (Promega, Madison, Wisconsin), 10 pmol of each primers, and 2 µl of the purified template. The amplification conditions were as described above avoiding the reverse transcription step.

C-PrM region and whole-genome nucleotide sequencing

For analyzing the C-PrM region, approximately 50 ng of the gel purified 654-bp PCR product was subjected to direct nucleotide sequencing reaction without cloning. For whole-genome analysis, larger sub-genomic fragments (approximately 2 kb; Fig. 1) were amplified and cloned in plasmid vectors. For this, the virus was isolated in C6/36 mosquito cell lines. Confluent cell monolayers in T-25 flasks were infected with 1:10 diluted dengue positive serum samples and further cultured in L-15 medium (Invitrogen, USA) supplemented with 2 % Fetal bovine serum (Invitrogen, USA) at 28 °C. The virus containing supernatants were collected on fifth day post-infection as described earlier [30] and stored in aliquots at −80 °C. Isolates obtained after a single passage in C6/36 cells were used for sequencing. 2 µl of the viral RNA was isolated as previously described and used in a single-step RT-PCR done with a high fidelity enzyme mix (Fidelitaq RT-PCR kit; USB, Cleveland, Ohio) as per kit protocols using the primer combinations as shown in Fig. 1. The amplification conditions were as described previously. These were then cloned using pGEM-T easy vector kit (Promega, Madison, WI) as per manufacturer's instructions, and plasmid DNA was isolated from the JM109 strain of E. coli transformed with the vector and cultured on ampicillin (50 µg/ml) containing selection medium. Six overlapping clones were made for each isolate to span the entire virus genome. The PCR products and the clones were sequenced bi-directionally using the Big-dye Terminator Cycle sequencing kit in an ABI 3730 Genetic Analyzer automated DNA sequencer (PE Applied Biosystems, Foster City, CA). Accuracy of the sequencing reads were ensured by selecting only the high quality base-calls as per the chromatogram and also by sequencing additional clones. The primers used for the genome walking approach are given in Table 2. The sequence contigs were assembled using CAP contig assembly program in the Bio-Edit 6.0.7 software [33].

Table 1 Primers used for Dengue viral RNA detection in the clinical samples and serotype identification by RT-PCR

Primer name	Sequence (5′ → 3′)	Location (with respect to the ref. sequence)	Reference sequence (GenBank accession no)	Amplicon size (bp)	Reference (publication)
D1F	TCAATATGCTGAACGCGCGAGAAACCG	132–159	NC_001477		[31]
DencomR2	GCNCCTTCDGMNGACATCC	783–765	NC_001477	654	[30]
nTS1	CTGGTTCCGTCTCAGTGATCCGGGGG	620–595	NC_001477	489	[31]
nTS2	AACGCCACAAGGGCCATGAACA	254–233	AY858096	123	
nTS3	TGCTGGTAACATCATCATGAGACAGAGCG	427–399	NC_001475	296	
nDen4	CTCTGTTGTCTTAAACAAGAGAGGTC	527–502	NC_002640	395	[32]

Fig. 1 Strategy adopted for whole-genome sequencing of dengue virus serotype 1. The primers used for the RT-PCR amplification of the six sub-genomic clones and for sequencing by genome walking approach are given as Table 2
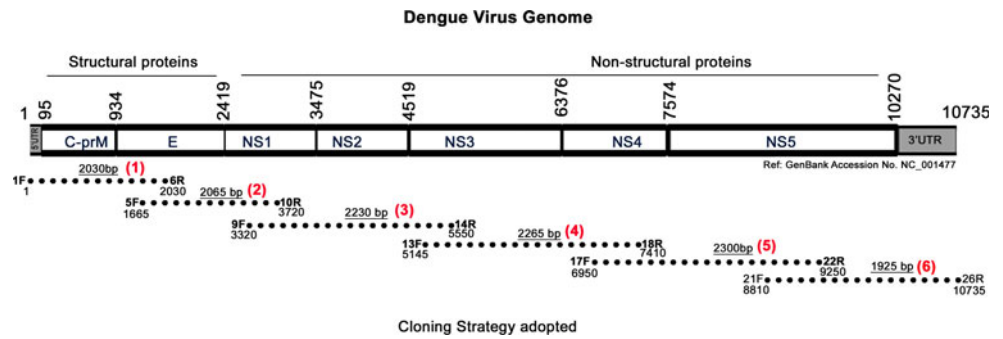
Table 2 Primers used in the study for whole-genome sequencing

Name	Forward primer sequence (5′–3′)	Position (with respect to NC_001477)	Name	Reverse primer sequence (5′–3′)	Position (with respect to NC_001477)
D1seq1F	tacgtggaccgacaagaacagtttcg	10–35	D1seq2R	cctgcttgctaactatcatgtgtgg	462–486
D1seq2F	ccacacatgatagttagcaagcagg	462–486	D1seq3R	caccgtgaatcctgggtgtc	820–839
D1seq3F	gagacacccaggattcacggtg	818–839	D1seq4R	gttcgtcgacacacaaagttcg	1,195–1,216
D1seq4F	gaactttgtgtgtcgacgaacg	1,196–1,217	D1seq5R	gctgtcttgaatgtgaccagca	1,639–1,660
D1seq5F	ctggtgacatttaagacagctcatg	1,641–1,665	D1seq6R	tctcaccaaaaggccggttct	2,039–2,058
D1seq6F	agaaccgcctttttggtgaga	2,039–2,058	D1seq7R	gtgacgaatgccacttccaca	2,460–2,482
D1seq7F	tgtggaagtggcatttttgtcac	2,460–2,482	D1seq8R	tggtcatcgggacattctggag	2,836–2,857
D1seq8F	ccaaacaccccagaatgc	2829–2,846	D1seq9R	ttccaacttgcctaggtgccatg	3,217–3,239
D1seq9F	ccatctcttaggaccacaacagtca	3,303–3,327	D1seq10R	acattggtctcatttaaaagtggc	3,708–3,732
D1seq10F	gccacttttaaaatgagaccaatgt	3,708–3,732	D1seq11R	tgggccggctagtggcacgtc	4,200–4,220
D1seq11F	tggcccctcaatgaaggaatcatggct	4,131–4,158	D1seq12R	gaggacagccccccctggtga	4,672–4,691
D1seq12F	tggcatgtcaccaggggagct	4,665–4,685	D1seq13R	ctcacggactatggctggaa	5,131–5,150
D1seq13F	ttccagccatagtccgtgag	5,131–5,150	D1seq14R	gagttccatgatctctcaggaa	5,536–5,557
D1seq14F	ttcctgagagatcatggaactc	5,536–5,557	D1seq15R	agggctgggataattccttctgg	6,021–6,043
D1seq15F	ccagaaggaattatcccagccct	6,021–6,043	D1seq16R	agcgtcaaatgctgtggaagttt	6,408–6,430
D1seq16F	tagggaaacttccacaacactt	6,405–6,426	D1seq17R	gccactgcatagagggtccagg	6,937–6,960
D1seq17F	cctggaccctctatgtcagtggc	6,937–6,958	D1seq18R	gccagtgtgatggattcgcaca	7,396–7,417
D1seq18F	tgtgcgaatccatcacactggc	7,396–7,417	D1seq19R	gccacctcttccacaaccgagg	7,808–7,829
D1seq19F	cctcggttgtggaagaggtggc	7,808–7,829	D1seq20R	gccacatgtcttgttccagcg	8,345–8,365
D1seq20F	caatggctcacaggaaaccaac	8,299–8,319	D1seq21R	aacacagctcctattgctgcg	8,786–8,806
D1seq21F	tggagcagtgttcgttgacgaa	8,797–8,818	D1seq22R	agtagggcatgttcgggttcca	9,238–9,259
D1seq22F	tggaacccgaacatgccctact	9,238–9,259	D1seq23R	gttcatcttggttgcggcatg	9,748-9,768
D1seq23F	catgccgcaaccaagatgaac	9,748–9,768	D1seq24R	gcctatcagggatcccacacca	10,104–10,124
D1seq24F	ccaccaacatacaagtggccata	10,138–10,160	D1seq25R	ggtgttgggcccccgctgctgcg	10,542–10,563
D1seq25F	aacaccaggggaagctgtaccc	10,558–10,579	D1seq26R	tgattcaacggcaccattccat	10,703–10,724

Sequence and phylogenetic analysis

ClustalW function of the Bio-Edit 6.0.7 software was used for the comparative analysis of nucleotide and amino acid sequences of Indian as well as closely related Indo-pacific isolates. Phylogenetic analysis was carried out employing MEGA 5.5 program [34]. The data sets for the analysis were selected from the National Center for Biotechnology Information (NCBI) GenBank database (http://www.ncbi.nlm.nih.gov) based on the previous studies employing Indian strains [20–23] and the Indo-Pacific strains [12, 15–17]. A data set of 158 sequences for the C-PrM region and a data set of 59 sequences for the complete coding region were used in the analyses. Sequences were aligned by Clustal W [35] and the nucleotide substitution model for each data set was identified by the Model (ML) function in the MEGA 5.5. Accordingly, K + G (Kimura-2 parameter [36] with Gamma distribution) settings for the C-PrM region data sets, and GTR + G + I (General time reversible model with Gamma distribution with Invariant sites)

settings for complete coding region sequence data sets were used, and maximum-likelihood analysis with 1,000 bootstrap replications was carried out for each data sets.

Selection pressure analysis

For detection of evidences for positive selection, the online adaptive evolution server http://www.datamonkey.org/ was used [37]. Three different codon-based likelihood methods—single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), and random effects likelihood (REL)—executed by the server were employed to estimate the ratio of non-synonymous to synonymous substitutions (dN/dS ratio) per codon sites. The SLAC and FEL are conservative methods for estimation of site-by-site substitution rates, whereas the REL is a less conservative method best suited for small data sets with low sequence divergence and allows synonymous rate variation. The analysis was run with default parameters that used a neighbor joining tree and a significance setting of P value/Bayes factor <0.1 for SLAC and FEL and a Bayes Factor of 50 for REL analysis. The TrN93 nucleotide substitution model was predicted by the server as the best substitution model for the data set used, and this was employed in the analysis.

Results

Clinical picture of the outbreak

76 patients (44 males and 32 females), who were mostly interns and nursing students, with a mean age of 27.07 years were suspected to have dengue infection during the outbreak. Going by the World Health Organization case definitions, 13 patients (17.1 %) were clinically diagnosed as dengue hemorrhagic fever (DHF), 2 patients (2.6 %) as dengue shock syndrome (DSS), and the remaining cases as Dengue fever. The major presenting features were fever (97.4 %), myalgia (84.2 %), periorbital pain (48.7 %), backache (46 %), giddiness (35.5 %), abdominal pain (27.6 %), polyarthralgia (21 %), and syncope (15.8 %). The average fever duration was 3.03 days. Major physical findings included petechiae (21.3 %), hepatomegaly (13.15 %), splenomegaly (7.9 %), hypotension (6.57 %), overt bleeding (2.6 %), and ascites (1.3 %). Myocarditis was diagnosed in 3 (3.9 %) and 2 (2.6 %) had pancreatitis. Leucopenia (<3,000/mm^3) was present in 38 (50 %) and 41 (53.9 %) had thrombocytopenia (<100,000/mm^3). There was no mortality.

Laboratory diagnosis by serology and RT-PCR

Among the 76 samples, IgM antibody was positive in 58 (76.3 %), IgG antibody was positive in 36 (47.3 %) and both were positive in 28 (36.8 %). In the RT-PCR, 39 (51.3 %) were positive for dengue virus nucleic acid as identified by the amplification of the 654-bp product. In the serotype-specific multiplex PCR, all the DENV nucleic acid positive samples amplified a 489-bp fragment indicating that they belonged to the serotype 1 of dengue virus. Subsequent sequence analysis of the amplified C-PrM region confirmed this observation.

C-PrM region and whole-genome nucleotide sequencing

Of the 39 samples positive for DENV 1 infection, partial C-PrM region of the 37 samples could be sequenced by direct sequencing approach (Supplementary Table 1). Sequencing of the two samples could not be carried out as the recovery of primary PCR product from these samples was very low. For whole-genome analysis, two isolates that represented the major viral population in the outbreak (RGCB585/2009 and RGCB592/2009) (Fig. 2) were sequenced. Apart from this, we also sequenced two isolates from our repository that were collected from Thiruvananthapuram from sporadic cases in the years 2007 and 2008 (RGCB294/2007 and RGCB419/2008) (Supplementary Table 1). The total length of the assembled sequences were 19 bp less than the NCBI NC_001477 DENV-1 reference sequence as the optimal PCR primer regions designed for the study did not include the sequences from the extreme end of the 5′- and 3′-NTR.

Sequence analysis

To identify whether the 2009 outbreak strains were previously circulating in India or were newly introduced, we initially analyzed a short 303-bp consensus coding region of the C-PrM protein (corresponding to the position 257–559 of the NCBI DENV-1 reference sequence NC_001477). This region was selected for analysis as the unpublished nucleotide sequences of the corresponding region of DENV-1 strains from different geographical parts of India were available in the GenBank. This 303-bp C-PrM nucleotide sequences when compared, was found to be identical in 32 of the 37 samples studied. The region had 96 % identity with the NCBI NC_001477; 96.96 % identity with 1956 Indian prototype (P23086; GenBank Accession No. EU626489) and 95.58 % identity with the Singapore 8114/93 prototype (GenBank Accession No. AY762084). Five samples (RGCB601, 606, 611, 629, 666) had a nucleotide substitution T511C and one among them (RGCB606) had an additional substitution of T357C. The latter substitution resulted in an amino acid change of I88T in the capsid protein.

Fig. 2 Phylogenetic analysis of the partial (303 bp) C-PrM region of the Indian and closely related strains of dengue virus serotype-1. Maximum-likelihood analysis was carried out with 1,000 bootstrap replications employing a Kimura-2 parameter correction with Gamma distribution settings. The scale bar represents the number of nucleotide substitutions per site. Boot–strap values more than 50 % are shown. GenBank accession number and name, place and year of isolation are indicated. Sequences obtained in the study are shown (*filled triangle*) and the outbreak strains are *underlined*. Sequences of Indian strains are indicated with *open diamond*
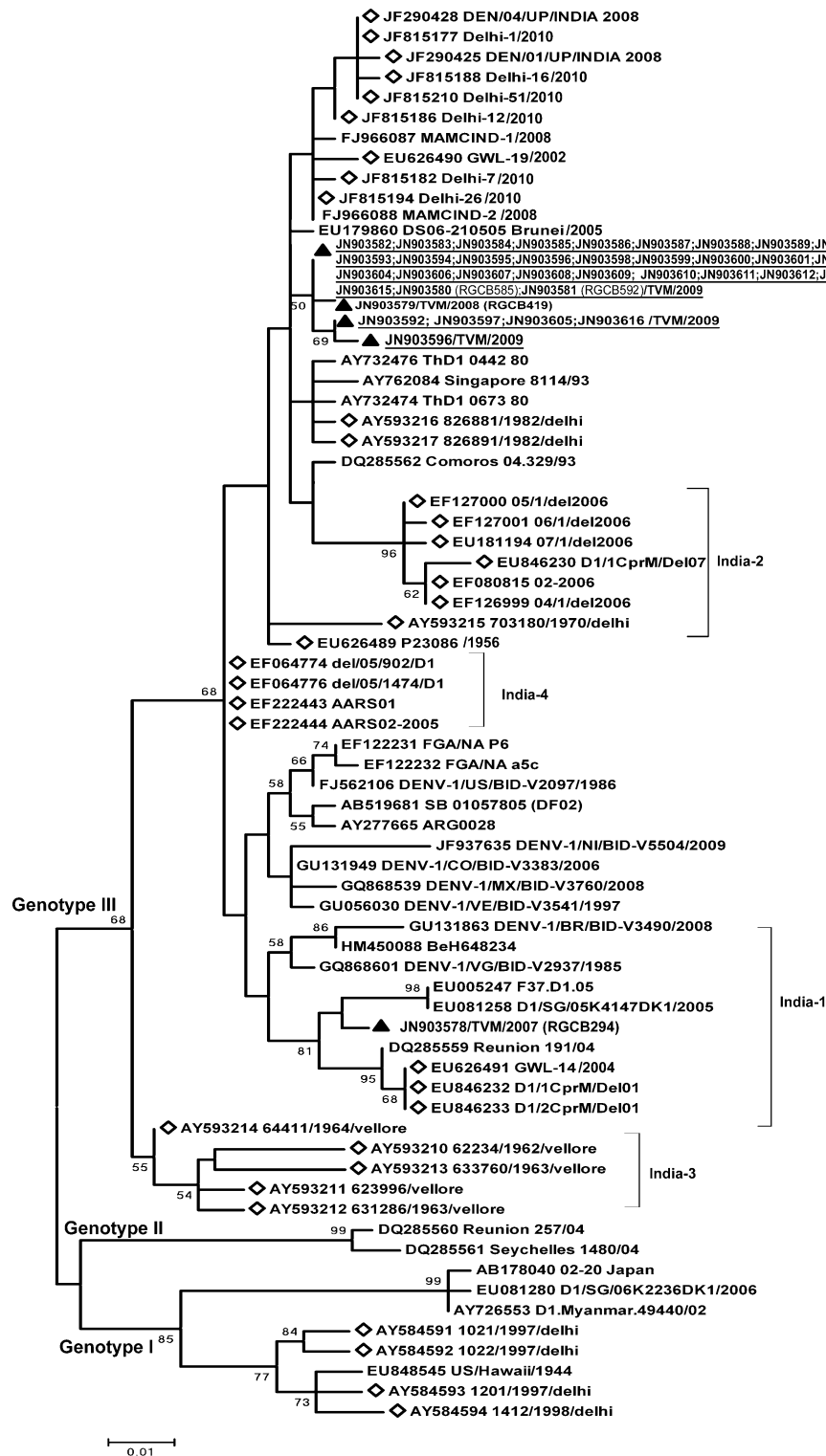
The full-length genome of the two strains from the outbreak RGCB585/2009 (GenBank accession no. JN903580); RGCB592/2009 (GenBank accession no. JN903581), and the 2008 strain RGCB419/2008 (GenBank Accession No. JN903579) were of 10,693 bp each, whereas that of

RGCB294/2007 (GenBank accession no. JN903578) was 10,714 bp. This difference in length was due to a 21-bp deletion (corresponding to the positions between 10,293 and 10,315) within the 3′-NTR of RGCB585/2009, RGCB592/2009, and RGCB419/2008 (Fig. 4). In online

basic local alignment search tool (BLAST) analysis, RGCB294/2007 showed a nucleotide level identity of 97 % with D1/SG/05K4147DK1/2005 (GenBank accession no. EU081258). The three strains RGCB585/2009, RGCB592/2009, and RGCB419/2008 showed an identity of 98 % with the Thailand isolates (ThD1_0442_80 and ThD1_0673_80) and 97.7 % identity with Brunei DS06-210505, Comoros 04.329/93, and Singapore 8114/93 isolates at the complete coding region nucleotide level. At the amino acid level, this translated to an identity of 98.2 % with the Brunei DS06-210505 and 99 % with the Singapore 8114/93 isolates. The three strains had 99.6 % similarity with the Singapore 8114/1993 isolate and 98.9 % similarity with the Brunei DS06-210505 isolate at the amino acid level. More detailed comparison of the sequences of these four isolates from Kerala and closely related Brunei, Thailand, Comoros, and Singapore strains revealed several non-synonymous substitutions leading to amino acid changes (Tables 3, 4). Many of them fell within the previously reported functional regions of the structural and non-structural proteins of the virus (Supplementary Fig. 1).

Phylogenetic analysis

The partial C-PrM region sequence (303 bp) was used in the initial phylogenetic analysis as previous studies from India have used this genomic portion [20, 21]. In the analysis, all the strains of this study clustered within the Genotype III clade of DENV-1 (Fig. 2). The isolate from 2007 (RGCB294/2007) grouped within the India-1 lineage [22, 23] designated as India-4 by Kukreti et al. [20], along with the North Indian strains from Delhi and Gwalior, and the Singapore isolate 05K4147DK. However, all the strains

sequenced from this outbreak were found to cluster as a separate clade with a significant bootstrap support. The 32 isolates with identical sequences clustered together (Fig. 2) representing the major viral population in the outbreak.

In the phylogenetic analysis using the full-length coding region sequences, the two major clades of American strains and the Indo-pacific strains were clearly evident within the Genotype III strains (Fig. 3). The two outbreak samples RGCB585/2009 and RGCB592/2009 and the 2008 isolate RGCB419/2008 clustered with the Brunei DS06-210505, Singapore 8114/93, Comoros 04.329/93, and two strains from Thailand (AY732474 and AY732476) in the Indo-pacific group. The RGCB294/2007, along with the Singapore 05K4147DK1 and RéUnion 191/04, formed a sub-clade within the American clade.

Selection pressure analysis

As nucleotide substitutions were observed in the isolates studied by whole-genome sequencing, we carried out selection pressure analysis across the coding region sequences. A single data set containing the complete coding region sequences of 12 strains that formed a distinct clade (RGCB294/2007, RGCB419/2008; RGCB585/2009; RGCB592/2009, Singapore 8114/93, ThD1_0442_80, ThD1_0673_80, Brunei DS06-210505, Comoros 04.329/93, Myanmar 40568/98, Myanmar 40553/96 sequences; Fig. 3) was used in the analysis. The initial alignment of the sequences was made with Singapore 8114/93 as the reference strain for the input data into the Datamonkey server. Among the three methods used, the mean dN/dS ratio obtained for SLAC method was 0.076, for FEL method was 0, and for REL was 0.106029. As per the

Table 3 Nucleotide changes in Kerala isolates with respect to Singapore 8114/93 reference strain

Isolate	Year of collection	Passage no. in C6/36 cells	GenBank accession no.	Number of Nucleotide changes											3′-NTR
				5′-NTR	Coding region (synonymous/non-synonymous)										
					Structural Protein		Non-structural Protein								
					C-PrM	E	NS1	NS2A	NS2B	NS3	NS4A	NS4B	NS5		
RGCB294/ 2007	2007	1	JN903578	0	34/2	73/7	61/7	21/4	11/1	67/3	27/3	40/5	158/12	14
RGCB419/ 2008	2008	1	JN903579	0	21/3	32/2	23/5	21/4	11/1	35/2	14/2	14/0	66/8	3
RGCB585/ 2009	2009	1	JN903580	0	23/3	34/3	22/5	22/5	12/2	35/2	15/1	13/0	68/10	4
RGCB592/ 2009	2009	1	JN903581	0	32/2	35/4	25/7	24/6	12/2	36/2	16/1	14/1	66/8	3

Table 4 Unique amino acid substitutions observed in the Kerala strains and the corresponding residues in the reference strain 8114/93, and the closely related DENV-1 isolates

Protein	Position		Isolates										
	Polypeptide	Protein	Singapore 8114/93	Comoros 04.329/93	ThD1 0442 80	ThD1 0673 80	DS06-210505	RGCB419/2008	RGCB585/2009	RGCB592/2009	RGCB294/2007	Re' Union 191/04	SG/05K4147DK1
C (Nt 95-436) Anchored capsid protein	2	2	N	.	.	.	.	.	D	.	.	.	.
	46	46	L	.	.	.	M	M	M	M	.	.	.
prM (Nt 437-934) Membrane glycoprotein precursor	129	15	S	.	.	.	.	G	.	.	.	.	.
	143	29	A	.	.	.	.	.	.	.	V	V	V
	232	118	K	R	R	R	R	R	R	R	R	.	.
	236	122	K	.	.	.	.	.	.	.	R	.	R
E (Nt 935-2,419) Envelope	409	129	I	.	.	.	.	.	.	.	V	.	V
	531	251	V	.	.	.	.	.	.	A	.	.	.
	577	297	T	V	V	V	V	V	V	V	M	M	M
	618	338	S	.	.	.	.	.	.	.	L	.	.
	619	339	T	.	.	.	.	.	.	.	I	.	I
	660	380	V	I	I	I	I	I	I	I	.	.	I
	671	391	W	.	.	.	.	.	.	.	R	.	.
	716	436	V	.	.	.	.	.	.	.	I	T	I
	752	472	S	.	.	.	N	.	N	N	.	.	.
	761	481	A	V	.	.	.	.	.	.	T	.	T
NS1 (Nt 2,420-3,475)	777	2	S	.	.	.	T	T	T	T	.	.	.
	873	98	T	I	I	I	.	.	.	.	A	A	A
	903	128	T	I	I	I	I	I	.	.	I	I	I
	962	187	A	.	.	.	.	.	.	V	V	.	.
	995	220	V	I	.	.	I	.	.	.	I	.	.
	1,002	227	K	R	R	R	R	R	R	R	R	.	.
	1,031	256	Y	.	.	.	.	.	.	.	H	.	.
	1,053	278	D	.	.	.	N	N	N	N	.	.	.
	1,054	279	L	.	.	.	.	.	.	.	V	.	.
	1,082	307	T	I	I	I	I	I	I	I	I	.	.
	1,099	324	K	R	.	.	R	R	R	R	R	R	R
	1,116	341	K	.	.	.	.	.	.	E	.	.	.
NS2A (Nt 3,476-4,129)	1,151	24	R	.	.	.	.	.	K	K	.	.	.
	1,241	114	V	.	.	.	A	A	A	A	A	A	A
	1,265	138	T	.	.	.	.	.	.	A	.	.	.
	1,295	168	M	I	I	I	I	I	I	I	I	.	.
	1,298	171	V	.	.	I	I	I	I	I	I	T	A
NS2B (Nt 4,130-4,519)	1,451	105	V	.	.	.	.	.	I	.	.	.	.

Table 4 continued

Protein	Position		Isolates										
	Polypeptide	Protein	Singapore 8114/93	Comoros 04.329/93	ThD1 0442 80	ThD1 0673 80	DS06-210505	RGCB419/2008	RGCB585/2009	RGCB592/2009	RGCB294/2007	Re' Union 191/04	SG/05K4147DK1
NS3	1,522	47	H	.	.	.	.	.	.	R	.	.	.	.
(Nt 4,520-6,377)	1,866	391	F	.	.	.	.	.	I	.	.	.	.	.
	1,906	431	K	.	.	.	.	.	.	.	E	.	.	.
	1,927	452	A	V	V	.	.	.	.	.	.	V	V	V
	1,949	474	I	.	.	V	.	.	.	.	.	V	V	V
	1,959	484	E	.	.	.	.	.	.	.	.	G	.	.
NS4	1,522	47	H	.	.	.	.	.	.	R	.	.	.	.
(Nt 6,378-7,573)	2,187	93	V	.	.	.	.	.	.	.	.	A	A	A
	2,235	141	L	.	.	.	.	.	S	.	.	.	.	.
	2,255	161	K	.	.	.	.	.	.	.	R	.	.	.
	2,263	169	V	.	.	.	.	.	.	.	.	A	A	A
	2,275	181	I	.	.	.	.	.	.	.	.	V	V	V
	2,278	184	H	.	.	.	.	.	.	.	.	R	R	R
	2,292	198	I	.	.	.	.	.	.	.	.	V	V	V
	2,491	397	G	.	.	.	.	.	.	.	.	S	.	.
NS5	2,494	1	G	.	.	.	.	.	D	.	.	.	.	.
(Nt 7,574-10,270)	2,516	23	S	.	.	.	.	.	P	.	.	.	.	.
	2,553	60	A	.	.	.	.	.	.	.	V	.	.	.
	2,616	123	L	.	.	I	I	.	I	I	.	.	.	.
	2,628	135	T	.	.	.	I	.	I	.	.	I	I	I
	2,763	270	I	.	.	.	.	.	.	M	.	V	V	V
	2,773	280	I	.	.	.	.	.	.	.	.	.	.	.
	2,858	365	P	.	.	.	.	.	.	.	.	A	A	A
	3,005	512	H	.	.	.	.	.	Y	Y	Y	.	.	.
	3,122	629	S	.	.	.	.	.	.	.	.	L	L	L
	3,128	635	T	N	N	N	N	N	N	.	N	N	N
	3,129	636	P	S	S	S	S	S	S	.	S	S	.	.
	3,133	640	E	.	.	.	.	.	.	.	.	K	K	K
	3,144	651	V	.	.	.	.	.	.	.	.	A	A	A
	3,162	669	T	.	.	.	.	.	.	.	.	I	I	I
	3,278	785	N	.	.	.	.	.	.	.	.	D	D	D
	3,327	834	D	.	.	.	.	.	.	.	.	E	E	E
	3,372	879	L	.	.	.	S	.	S	S	S	.	.	.
	3,375	882	M	.	.	.	.	.	T	.	.	.	.	.
	3,377	884	S	.	.	.	.	.	.	.	.	.	.	T

The sequences from the study are emphasized in bold and the Domain III (the major immunogenic domain in the envelope protein) coding region is italicized

Fig. 3 Phylogenetic analysis using the complete protein coding nucleotide sequences. Maximum-Likelihood analysis was carried out with 1,000 bootstrap replications employing a general time reversible model (GTR) with gamma (G) distributed with Invariant sites (I) settings. The *scale bar* represents the number of nucleotide substitutions per site. Bootstrap values more than 50 % are shown. GenBank accession number and strain names are given for each strain. Sequences obtained in the study are shown (*filled triangle*)
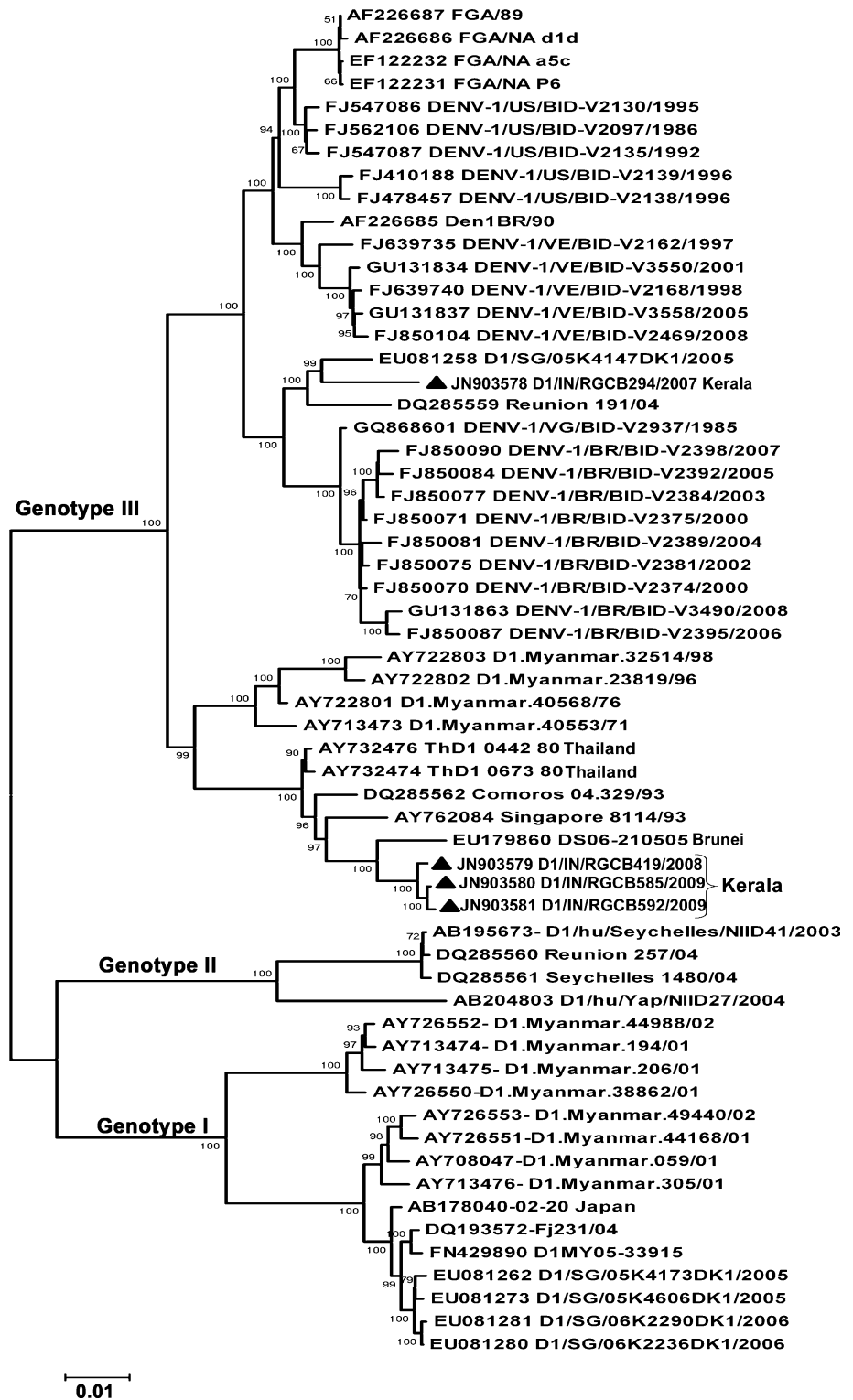
statistical significance criteria set for the program (*P* value/ Bayes factor <0.1 for SLAC and FEL and a Bayes Factor of 50 for REL), none of these methods identified sites with positive selection that enhances the evolutionary fitness of the viral strains studied.

Discussion

We initiated this study to characterize the viral strains from a dengue outbreak which was unique in being a non-seasonal outbreak. Seasonality of Dengue infection in India

Fig. 4 Comparison of the 3′-NTR of the Kerala isolates and closely related Indo-pacific DENV-1 isolates. Conserved residues are indicated by *dots*
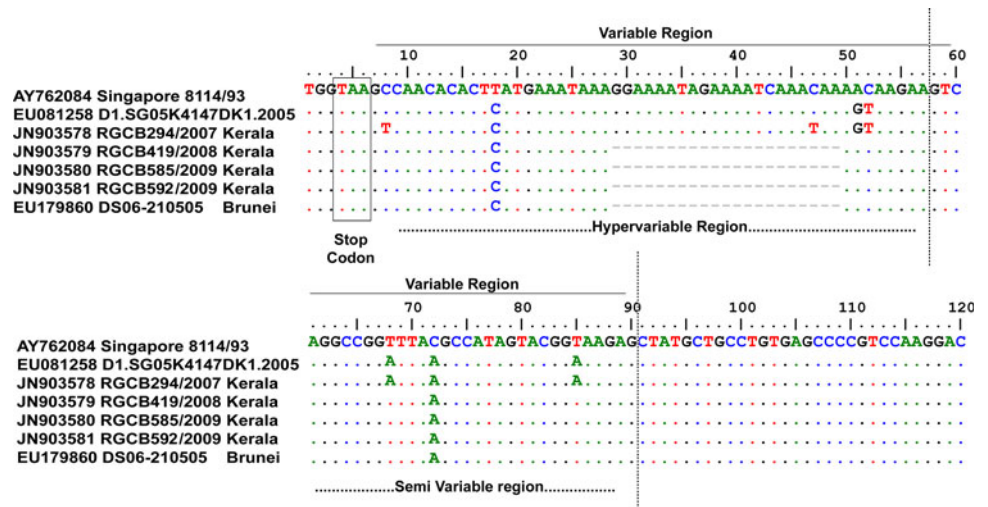
has been previously studied [19, 38]. Over the years, a strong association between the Monsoon and the dengue occurrence in the Indian subcontinent has been observed, with an increased incidence during the Monsoon and immediate post-monsoon periods (June–October) [18, 19, 28, 29]. A few earlier studies have reported the occurrence of dengue outbreaks in dry season of March–May [39, 40]. In Kerala, this period is characterized by the occurrence of pre-monsoon showers with irregular dry and wet spells [41] that favors localized *Aedes* mosquito breeding activity and resurgence of disease incidence in Dengue endemic areas of the state. The clinical severity of the present outbreak was moderate with only 13 cases of DHF (17.1 %) and two cases of DSS (2.6 %) and no mortality. The local authorities could identify the breeding source of mosquitoes in the present outbreak as the waterlogged areas in a building construction site in the medical college campus, and could contain the spread of the disease successfully with stringent vector control measures.

The major observation in this study was the genetic distinctness of the DENV-1 strains involved in the outbreak from the previously reported Indian DENV-1 strains. Even though the disease outbreak under study occurred only in 2009, closely related strains of the DENV-1 that caused the outbreak were circulating in the region in earlier years. This was indicated by the close similarity of the two strains of the outbreak (RGCB585 and 592) with the 2008 isolate (RGCB419) from Thiruvananthapuram. These three strains were identical with the presence of four mutations—L46M in the capsid, D278N in the NS1, L123I and L879S in the NS5 protein coding regions (Table 4). Also, they shared a 21-bp deletion in the 3′-NTR (Fig. 4). The possibility of the observed genetic variation in the strains causing this outbreak could be due to extensive local evolution of this highly mutating virus [3, 4] accumulating nucleotide changes during sporadic infections in this dengue endemic

area, or might be due to presence of strains introduced from elsewhere. The latter possibility of exotic introduction into the state is highly likely as supported by the finding that the three strains (RGCB419, RGCB585, and RGCB592) from Kerala shared features with the 2005 Brunei isolate DS06-210505 [15] and the ThD1_0442_80, ThD1_0673_80, and Comoros 04.329/93 strains. With the Brunei isolate, this pertained to the common presence of unique amino acid substitutions such as the L46M in capsid, D278N in the NS1, L123I, and L879S in the NS5 protein coding regions (Table 4), and also the 21-bp deletion in the 3′-NTR (Fig. 4). And with the Thailand and Comoros strains, this was indicated by the uniform presence of the amino acid substitutions K118R in the PrM, T297V, and V380I in the envelope, K227R and T307I in the NS1, M168I in the NS2A, T635N and P636S in the NS5 proteins (Table 4 and Supplementary Fig. 1). The phylogenetic clustering of these sequences with a high bootstrap support (100 %) (Fig. 3) also supported the conclusion. The observation is not surprising as earlier studies have shown circulation of dengue viral strains with significant genetic identity in countries of the Indo-pacific rim indicating frequent cross-transmission of the viral strains in the region [12, 20, 22, 23].

Involvement of strains genetically similar to the ones obtained in this study could not be identified in previous dengue outbreaks in India based on the available literature. Nevertheless, as the deletion in the 3′-NTR was a unique feature, we did an online BLAST search for strains showing the 21-bp deletion in the 3′-NTR. We could identify one previous strain from India collected from an outbreak in Delhi during 2006 (NCBI GenBank accession no. EU418660; unpublished) that had this deletion. This indicates the presence of strains with the 3′-NTR deletion in the country. However, as sequences of other genomic regions of this strain are not available for comparison, it

was not possible to identify whether the strain is related to the three strains (RGCB419, RGCB585, and RGCB592) from Kerala. The possibility seems unlikely as indicated by our phylogenetic analysis with the C-PrM sequences wherein the DENV-1 strains of the 2006 Delhi outbreak used in the previous studies [20, 21] cluster separately as a distinct lineage (India-2) from the 2008 and 2009 Kerala strains (Fig. 2). It might also be possible that even if the strains with 3′-NTR deletion was circulating in 2006 Delhi outbreak, it was only a minor population that was not picked up in these studies.

As evidenced by the phylogenetic analysis (Fig. 2), the strains that caused the present outbreak are genetically distinct from the India-2 lineage and the previously circulating strains in the region. However, it might not be prudent to designate them as an independent lineage as the bootstrap support for this clustering was not highly significant (only 50 %); the analysis used only a smaller (303 bp) sequence of the C-PrM; and the sequence divergence of this clade with the rest of the closely related Indian strains was low (1.5 %). Further analysis using complete genomic sequences of virus isolates from other regions are essential for clearly deciphering the genetic relatedness of the outbreak strains, RGCB585 and RGCB592 within the Indian DENV-1 strains.

The 21-bp deletion observed in the 3′-NTR is important as the region play a major role in replication of the Dengue genomic RNA [42]. The 3′-NTR has a variable region and a conserved region in dengue virus. The variable region is represented by the 84 nucleotides immediately following the stop codon [42–44]. Within this variable region, two regions viz. a hyper-variable region (HVR) and a semi-variable region (SVR) have been delineated. Nucleotide changes in the variable region can affect the efficiency of dengue virus replication [45]. The 21-bp nucleotide deletion observed in three of the Kerala strains (RGCB419, RGCB585, and RGCB592) was within the HVR region of the 3′-NTR (Fig. 4). The same deletion has also been observed in the Brunei strain [15]. An earlier study [45] has shown that among the 45 nucleotides of the HVR, a 26 nucleotides sequence alone is sufficient for efficient replication of the virus. The deletion observed in these strains was located downstream this region (Fig. 4) implying that the deletion may not affect the viral replication kinetics.

In the evolutionary dynamics of Dengue, genotype level changes are considered significant and are conventionally attributed to change the disease profile [7]. However, a few studies have speculated the role of minor genetic changes and lineage diversification, and have implicated them as the reasons for the observed increase in severity of recent DENV1 cases in the country [20, 23]. This highlights the importance of small-scale variations in the viral genome as

observed in this study. The results of the study support the progressively changing molecular epidemiology of dengue in India, and also point to the role of exotic introductions of viral strains in this process. The sequences generated would serve as a reference for future studies of the circulating DENV-1 strains in South Asia, especially in the Indian subcontinent.

References

1. J.G. Rigau-Perez, G.G. Clark, D.J. Gubler, P. Reiter, E.J. Sanders, A.V. Vorndam, Lancet **352**, 971–977 (1998)
2. S. Green, A. Rothman, Curr. Opin. Inf. Dis. **19**, 429–436 (2010)
3. J.W. Drake, Proc. Natl. Acad. Sci. **90**, 4171–4175 (1993)
4. E.C. Holmes, S.S. Burch, Trends Microbiol. **8**, 74–77 (2000)
5. E.C. Holmes, S.S. Twiddy, Infect. Genet. Evol. **3**, 19–28 (2003)
6. S.C. Weaver, N. Vasilakis, Infect. Genet. Evol. **9**, 523–540 (2009)
7. R. Rico-Hesse, L.M. Harrison, R.A. Salas, D. Tovar, A. Nisalak, C. Ramos, J. Boshell, M.T. de Mesa, R.M. Nogueira, A.T. da Rosa, Virology **230**, 244–251 (1997)
8. W.B. Messer, D.J. Gubler, E. Harris, K. Sivananthan, A.M. de Silva, Emerg. Infect. Dis. **9**, 800–809 (2003)
9. D.W. Vaughn, S. Green, S. Kalayanarooj, B.L. Innis, S. Nimmannitya, S. Suntayakorn, T.P. Endy, B. Raengsakulrach, A.L. Rothman, F.A. Ennis, A. Nisalak, J. Infect. Dis. **181**, 2–9 (2000)
10. R. Cologna, P.M. Armstrong, R. Rico-Hesse, J. Virol. **79**, 853–859 (2005)
11. C.Y. Huang, S. Butrapet, K.R. Tsuchiya, N. Bhamarapravati, D.J. Gubler, R.M. Kinney, J. Virol. **77**, 11436–11447 (2003)
12. M.J. Schreiber, E.C. Holmes, S.H. Ong, H.S. Soh, W. Liu, L. Tanner, P.P. Aw, H.C. Tan, L.C. Ng, Y.S. Leo, J.G. Low, A. Ong, E.E. Ooi, S.G. Vasudevan, M.L. Hibberd, J. Virol. **83**, 4163–4173 (2009)
13. M.G. Guzman, S.B. Halstead, P. Artsob, P. Buchy, J. Farrar, D.J. Gubler, E. Hunsperger, A. Kroeger, H.S. Margolis, E. Martinez, M.B. Nathan, J.L. Pelegrino, C. Simmons, S. Yoksan, R.W. Peeling, Nat. Rev. Microbiol. **8**, S7–S16 (2010)
14. G. Aviles, J. Meissner, R. Mantovani, S. St Jeor, Virus Res. **98**, 75–82 (2003)
15. O. Osman, M.Y. Fong, S.D. Sekaran, J. Gen. Virol. **90**, 678–686 (2009)
16. Y. Tang, P. Rodpradit, P. Chinnawirotpisan, M.P. Mammen Jr, T. Li, J.A. Lynch, R. Putnak, C. Zhang, Am. J. Trop. Med. Hyg. **83**, 1156–1165 (2010)
17. V. Duong, C. Simmons, L. Gavotte, A. Viari, S. Ong, N. Chantha, N.J. Lennon, B.W. Birren, S. Vong, J.J. Farrar, M.R. Henn, V. Deubel, R. Frutos, P. Buchy, Infect. Genet. Evol. (2011). doi:10.1016/j.meegid.2011.06.019
18. U.C. Chaturvedi, R. Nagar, J. Biosci. **33**, 429–441 (2008)
19. A. Chakravarti, R. Arora, C. Luxemburger, Trans. R. Soc. Trop. Med. Hyg. (2012). doi:10.1016/j.trstmh.2011.12.007
20. H. Kukreti, P.K. Dash, M. Parida, A. Chaudhary, P. Saxena, R.S. Rautela, V. Mittal, M. Chhabra, D. Bhattacharya, S. Lal, P.V. Rao, A. Rai, Virol. J. **6**, 1 (2009)

21. H. Kukreti, A. Chaudhary, R.S. Rautela, R. Anand, V. Mittal, M. Chhabra, D. Bhattacharya, S. Lal, A. Rai, Int. J. Infect. Dis. **12**, 542–549 (2008)

22. C. Domingo, G. Palacios, O. Jabado, N. Reyes, M. Niedrig, J. Gascon, M. Cabrerizo, W.I. Lipkin, A. Tenorio, J. Clin. Microbiol. **44**, 1519–1529 (2006)

23. J.A. Patil, S. Cherian, A.M. Walimbe, B.R. Patil, P.S. Sathe, P.S. Shah, D. Cecilia, Infect. Genet. Evol. **11**, 1443–1448 (2011)

24. D. Cecilia, M.B. Kakade, A.B. Bhagat, J. Vallentyne, A. Singh, J.A. Patil, S.M. Todkar, S.B. Varghese, P.S. Shah, Virol. J. **8**, 46 (2011)

25. S. Sharma, P.K. Dash, S. Agarwal, J. Shukla, M.M. Parida, P.V. Rao, J. Gen. Virol. **92**, 1595–1600 (2011)

26. R.M. Myers, M.J. Varkey, R. Reuben, E.S. Jesudass, B. Benjamin, Am. J. Public Health **61**, 1379–1391 (1971)

27. M. Kurukumbi, J.P. Wali, S. Broor, P. Aggarwal, P. Seth, R. Handa, L. Dhar, M. Vajapayee, Indian J. Med. Sci. **55**, 149–156 (2001)

28. B.K. Tyagi, J. Hiriyan, P. Philip Samuel, S.C. Tewari, R. Paramasivan, ICMR Bull. **36**, 13–30 (2006)

29. A. Kumar, C.R. Rao, V. Pandit, S. Shetty, C. Bammigatti, C.M. Samarasinghe, Indian J Commun Med. **35**, 386–390 (2010)

30. M. Anoop, A. Issac, T. Mathew, S. Philip, N.A. Kareem, R. Unnikrishnan, E. Sreekumar, Indian J. Exp. Biol. **48**, 849–857 (2010)

31. R.S. Lanciotti, C.H. Calisher, D.J. Gubler, G.J. Chang, A.V. Vorndam, J. Clin. Microbiol. **30**, 545–551 (1992)

32. E. Harris, T.G. Roberts, L. Smith, J. Selle, L.D. Kramer, S. Valle, E. Sandoval, A. Balmaseda, J. Clin. Microbiol. **36**, 2634–2639 (1998)

33. T.A. Hall, Nucl. Acids Symp. Ser. **41**, 95–98 (1999)

34. K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, Mol. Biol. Evol. **28**, 2731–2739 (2011)

35. J.D. Thompson, D.G. Higgins, T.J. Gibson, Nucleic Acids Res. **22**, 4673–4680 (1994)

36. M. Kimura, J. Mol. Evol. **16**, 111–120 (1980)

37. L. Sergei, P. Kosakovsky, S.D.W. Frost, Bioinformatics **21**, 2531–2533 (2005)

38. A. Chakravarti, R. Kumaria, Virol J. **2**, 32 (2005)

39. G.S. Chouhan, F.M. Rodrigues, B.H. Shaikh, M.A. Ilkal, S.S. Khangaro, K.N. Mathur, K.R. Joshi, N.K. Vaidhye, Indian J. Med. Res. **91**, 414–418 (1990)

40. A.R. Risbud, S.M. Mehendale, G.D. Joshi, K. Banerjee, Indian J. Virol. **7**, 120–127 (1991)

41. C.U. Warrier, M. Praveen Babu, P. Manjula, K.T. Velayudhan, A. Shahul Hameed, K. Vasu Curr, Science **98**, 1487–1495 (2010)

42. D.E. Alvarez, A.L. De Lella Ezcurra, S. Fucito, A.V. Gamarnik, Virology **339**, 200–212 (2005)

43. V. Proutski, E.A. Gould, E.C. Holmes, Nucleic Acids Res. **25**, 1194–1202 (1997)

44. A.C. Shurtleff, D.W. Beasley, J.J. Chen, H. Ni, M.T. Suderman, H. Wang, R. Xu, E. Wang, S.C. Weaver, D.M. Watts, K.L. Russell, A.D. Barrett, Virology **281**, 75–87 (2001)

45. S. Tajima, Y. Nukui, T. Takasaki, I. Kurane, J. Gen. Virol. **88**, 2214–2222 (2007)