Infection, Genetics and Evolution 12 (2012) 1759–1769

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Influence of evolutionary events on the Indian subcontinent on the phylogeography of dengue type 3 and 4 viruses

J.A. Patil [a], S. Cherian [b], A.M. Walimbe [b], A. Bhagat [a], J. Vallentyne [a], M. Kakade [a], P.S. Shah [a], D. Cecilia [a,*]

[a] Dengue Group, National Institute of Virology, 20-A, Dr. Ambedkar Road, P.O. Box No. 11, Pune 411001, Maharashtra State, India
[b] Bioinformatics Group, National Institute of Virology, 20-A, Dr. Ambedkar Road, P.O. Box No. 11, Pune 411001, India

ARTICLE INFO

Article history:
Received 24 February 2012
Received in revised form 12 July 2012
Accepted 15 July 2012
Available online 6 August 2012

Keywords:
DENV-3
DENV-4
India
Recombination
Evolution
Phylogeography

ABSTRACT

During 1960–80 dengue disease profile in India was mild despite circulation of all four serotypes of dengue virus (DENV). Increase in disease severity with a concomitant change in the population of DENV-1 and 2 have been reported since then. To determine population dynamics of DENV-3 and 4, the envelope (E) gene sequence was determined for 16 Indian isolates of DENV-3 and 11 of DENV-4 and analyzed together with 97 DENV-3 and 43 DENV-4 global sequences.

All Indian DENV-3 isolates belonged to genotype III, lineages C, D, E and F. Lineage F was newly identified and represented non-circulating viruses. Three non-conservative amino acid changes in domain I, II & III were identified during the transition from lineages F/E, associated with mild disease, to A–D, associated with severe disease. For DENV-4, the current viruses clustered in genotype I, lineage C, whilst the isolates from 1960s formed the new genotype V. A 1979 Indian isolate of DENV-4 was found to be an inter-genotypic recombinant of Sri Lankan isolate (1978) of genotype I and Indian isolate (1961) of genotype V. The rates of nucleotide substitution and time to the most recent common ancestor (tMRCA) estimated for DENV-3 (1782–1934) and DENV-4 (1719–1931) were similar to earlier reports. However, the divergence time for genotype III of DENV-3, 1938–1963, was a more accurate estimate with the inclusion of Indian isolates from the 1960s. By phylogeographical analysis it was revealed that DENV-3 GIII viruses emerged from India and evolved through Sri Lanka whilst DENV-4 emerged and dispersed from India.

The present study demonstrates the crucial role that India/Sri Lanka have played in the evolution and dispersion of the major genotypes, GIII of DENV-3 and GI of DENV-4 which are more virulent and show higher dissemination potential.

1. Introduction

During the last two decades changes in the circulating serotype/genotype of DENV have been reported from many parts of the world (Hang et al., 2010; Zhang et al., 2005). Some of these changes have been associated with increase in disease severity (Rico-Hesse, 2010). Similar observations have been reported from India; compared to the mild disease profile in the 1960s (Meyers et al., 1970) the number of severe cases has increased since the late 1980s, the first outbreak with Dengue hemorrhagic fever (DHF) cases being reported in 1989 (Pushpa et al., 1998). Over the past two decades, regular epidemics of Dengue fever (DF) and DHF have been reported in Sri Lanka, India, the Maldives Islands, Bangladesh, and Pakistan (Messer et al., 2002, 2003; Rahman et al., 2002).

In India change in disease severity was suggested to be a result of change in the genotype (G) of DENV-2 and a change in the lineage of DENV-1 on the basis of the E gene sequence (Kumar et al., 2010; Patil et al., 2011). Similar findings were reported on the basis of shorter sequences in the Core-prM region (Dash et al., 2004 and Kukreti et al., 2009). Emergence of DENV-3 (GIII) was reported in the 2005 outbreak in Northern India (Dash et al., 2006; Gupta et al., 2006). The same genotype, GIII was implicated in causing DHF outbreaks in the Americas (Palaez et al., 2004; Passos et al., 2004; Pan American Health Organization, 2007). DENV-4, which was not detected during the last three decades in India, was reported in Delhi in 2003 (Dar et al., 2006), Hyderabad in 2007 (Dash et al., 2011) and Pune in 2009 (Cecilia et al., 2011). The current DENV-4 isolates belonged to GI (Cecilia et al., 2011). A similar re-emergence of DENV-4, GI in Brazil in 2008 after three decades was reported by Melo et al. (2009).

The present study attempts to throw light on the evolution and dispersion of DENV-3 and 4 in India by sequencing and phylogeographic analyses of a relatively large number of Indian isolates from 1960s and the present times. The period of 1970–2000 has insufficient representation due to non-availability of isolates. The National Institute of Virology was actively involved in the

* Corresponding author. Tel.: +91 20 2600219; fax: +91 20 26122669.
E-mail addresses: ceciliadayaraj@gmail.com, cecilia.dayaraj@gmail.com (D. Cecilia).

1567-1348/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2012.07.009

investigation of dengue outbreaks since the 1950s when the virus first made its appearance in India (Carey et al., 1964) parallel with the first isolations in the world (Kimura and Hotta, 1943; Sabin, 1952). Therefore the unique set of viruses included in the present study represents a region and time that have not been sampled before. The data generated was thus expected to fill some of the lacunae in the understanding of dengue dispersion, evolution and phylogeography.

2. Materials and methods

2.1. Viruses

The envelope (E) genes of 16 DENV-3 isolates and 11 DENV-4 isolates from different parts of India were sequenced as part of this study (Tables 1A and 1B) and analyzed with global sequences of 97 DENV-3 and 43 DENV-4 isolates, retrieved from GenBank. The isolates obtained before 2000 were procured from the Virus Repository of the National Institute of Virology, Pune, at passage level 5 in infant mouse brain. After 2000, the isolates were obtained using C6/36 cells and sequenced at second passage level.

2.2. Sequencing

The viral RNA was extracted from 140 μl of infected C6/36 culture supernatant or 10% infected mouse brain suspension using QIAamp viral RNA mini kit (Qiagen) in accordance with the manufacturer's instructions. cDNAs were prepared using 10 μl RNA and Avian Moloney virus reverse transcriptase (Promega). The E gene was amplified using primers described in Supplementary Tables S1 and S2 using Taq polymerase (Invitrogen). PCR products were gel-purified using QIAquick gel extraction kit (Qiagen). The purified PCR products were sequenced using BigDye terminator cycle sequencing ready reaction kit (Applied Biosystems) on an automatic sequencer (ABI PRISM Genetic Analyzer 3100; Applied Biosystems).

2.3. Phylogenetic, molecular clock and phylogeography analysis

Clustal W implemented in Mega v. 3.1 was used for multiple sequence alignment. The percentage nucleotide identity and percentage amino acid identity values were calculated as pairwise *p* distances. Simultaneous estimation of phylogeny, rate of nucleotide substitution and divergence times (tMRCA) of DENV-3 and DENV-4 strains of Indian origin and also of the different genotypes was carried out using the Bayesian MCMC approach as implemented in BEAST 1.5.3 (Drummond and Rambaut, 2007). Recombinant sequences detected by the protocols mentioned above, were excluded from molecular clock analysis. The best-fit model of nucleotide substitution was selected by using Akaike Information Criterion (AIC) as implemented in MODELTEST 3.7 (Posada and Crandall, 1998). The GTR + G + I model (general time-reversible model with gamma-distributed rates of variation among sites and a proportion of invariable sites) was found to be the best-fit model for DENV-3 and DENV-4 datasets. We employed both strict and relaxed (uncorrelated exponential and uncorrelated lognormal) clock (Drummond et al., 2006) with different demographic models (constant size, exponential growth, logistic growth and expansion growth). The Bayesian skyline (BSL) plot was also constructed to study the genetic diversity of the population. Three independent MCMC analyses, each for thirty million steps, were performed for each combination of branch rate and demographic model, as well as for BSL and combined with a burn-in value set to 10% generations using Log Combiner program (implemented in BEAST). The convergence of the chain was evaluated by using Tracer 1.5 (Drummond and Rambaut, 2007). The effective sample size (ESS) values of >200 indicated sufficient level of sampling. The posterior probability and Bayes factor based on the marginal likelihoods of the models was used to choose the most suitable model for the data (Suchard et al., 2001). The maximum clade credibility tree was generated by using Tree Annotator program (available in BEAST), and Fig Tree 1.2.3 (http://tree.bio.ed.ac.uk/) was used for visualization of the annotated tree. The 95% HPD intervals were used to ascertain the uncertainty in the parameter estimates.

In addition, the spatial information of the sequences was used to infer the most probable ancestral geographic region/country at different internal nodes of the tree by fitting a standard continuous-time Markov chain (CTMC) model with the Bayesian stochastic search variable selection (BSSVS) in BEAST (Lemey et al., 2009). The ISO 3166-1 alpha 2 code was used to define the country of origin for isolates from Southeast Asian countries. The origin of isolates from the other parts of the world was defined by geographical region.

The MCMC tree obtained under the CTMC model was used as an input in SPREAD 1.0.3 (Bielejec et al., 2011) to visualize and analyze the dispersion pathways. The Bayes factor (BF) test available in SPREAD was used to identify well-supported rates of transitions between the different geo-regions.

2.4. Recombination analysis

Aligned sequences were submitted to Genetic algorithm for detection of recombination (GARD) (Sugiura, 1978; Kosakovsky-Pond

Table 1A
Indian DENV-3 isolates sequenced in this study.

No.	Strain	Year	Location	Passage level	Accession no.
1	NIV_664481_1966	1966	Chennai	P-19	JQ686068
2	NIV_664482_1966	1966/M[a]	Vellore	P-22	JQ686069
3	NIV_058760_2005	2005	Pune	P-02	JQ686078
4	NIV_059826_2005	2005	Pune	P-02	JQ686077
5	NIV_09401_2009	2009	Kerala	P-02	JQ686083
6	NIV_09509_2009	2009	Kerala	P-02	JQ686082
7	NIV_0910570_2009	2009	Sholapur/Pune	P-02	JQ686071
8	NIV_0920520_2009	2009	Pune	P-02	JQ686073
9	NIV_0948350_2009	2009	Pune	P-02	JQ686072
10	NIV_0948359_2009	2009	Pune	P-02	JQ686081
11	NIV_01016524_2010	2010	Pune	P-02	JQ686070
12	NIV_01029580_2010	2010	Pune	P-02	JQ686079
13	NIV_01029582_2010	2010	Pune	P-02	JQ686080
14	NIV_01030355_2010	2010	Pune	P-02	JQ686076
15	NIV_01030814_2010	2010	Pune	P-02	JQ686075
16	NIV_01023422_2010	2010	Pune	P-02	JQ686074

[a] Mosquito isolate.

Table 1B
Indian DENV-4 isolates sequenced in this study.

No.	Strain	Year	Location	Passage level	Accession no.
1	NIV_611319_1961	1961	Vellore	P-28	JQ686059
2	NIV_62231_1962	1962/M[a]	Vellore	P-11	JQ686066
3	NIV_62235_1962	1962/M[a]	Vellore	P-13	JQ686065
4	NIV_624000_1962	1962	Vellore	P-23	JQ686064
5	NIV_631302_1963	1963	Vellore	P-10	JQ686057
6	NIV_64431_1964	1964	Vellore	P-07	JQ686060
7	NIV_654021-2_1965	1965/M[a]	Maharashtra	P-03	JQ686062
8	NIV_654129-1_1965	1965/M[a]	Maharashtra	P-08	JQ686063
9	NIV_654129-2_1965	1965/M[a]	Maharashtra	P-03	JQ686061
10	NIV_793679_1979	1979	Not known	P-05	JQ686058
11	NIV_0952326_2009	2009	Kerala	P-02	JQ686067

[a] Mosquito isolate.

and Frost, 2005). Using break points obtained in GARD analysis, Neighbor joining phylogenetic trees were constructed to confirm the recombination event. An isolate was considered as a recombinant if a shift in its position was observed in the phylogenetic tree with good bootstrap support.

Putative recombinant sequences were re-confirmed using SIMPLOT version 3.2 (Lole et al., 1999). The putative recombinant sequence was queried against two potential parental sequences. A sliding window of 250 nucleotides was moved in steps of 5 nucleotides at a time and the resulting similarity values were plotted along the E gene sequence. Recombination was identified when conflicting E gene sequence profiles appeared, suggesting acquisition of sequences from a different parental genotype. Bootscan analyses which utilized the bootstrapping procedures of Salminen et al. (1995) and Worobey and Holmes (1999) were performed using the maximum-likelihood method with 100 re-samplings. Bootstrap values of 70% were used to indicate robust support for the topologies.

2.5. Selection pressure analysis

To identify the existence of positive selection pressure at individual codon sites in the E gene of DENV-3 and DENV-4, three likelihood procedures were used: SLAC, FEL and REL methods (Kosakovsky-Pond & Frost, 2005). The analysis was carried out using the Datamonkey server (http://www.datamonkey.org). Two datasets were used for DENV-3: one contained only Indian sequences ($n = 20$) while the other included all G III sequences ($n = 60$). For DENV-4 also two datasets were considered: one consisted of only Indian isolates ($n = 14$) and other consisted of all isolates of GI and GV ($n = 24$).

3. Results

3.1. Phylogenetic analysis

DENV-3: The MCMC tree revealed that the 95 DENV-3 isolates were distributed into five genotypes, I–V, with six lineages (A–F) in GIII (Fig. 1). The distribution of isolates in genotypes I, II, IV and V was similar to that reported earlier (Aruajo et al., 2009). All Indian DENV-3 isolates grouped into GIII. The Indian isolates of 1966, including one mosquito (NIV_664482_1966) and one human (NIV_664481_1966), were the oldest representative isolates of GIII, and formed a new lineage, lineage F. Lineage E, which was not a tight cluster, consisted of isolates from Sri Lanka (pre-DHF era), India (1984) and Samoa (1986). Lineage D included one Indian isolate of 2009, from Kerala, along with isolates from Sri Lanka (1993–2000, post-DHF) and Taiwan. Lineage C included Indian isolates of 2003–2010 along with isolates from Cambodia, Bhutan, Saudi Arabia, Abidjan and Tanzania. Lineage B consisted of isolates from Sri Lanka (post-DHF era) and Somalia. Lineage A was represented by American isolates. Nucleotide (nt) and amino acid (aa) sequence comparison of GIII lineages revealed low diversity within (1–1.4% nt; 0.3–1% aa) and between lineages, A–E (1.4–3.3% nt, 0.5–1% aa). In contrast Lineage F showed high intra-lineage (3% nt and 5.4% aa) and inter-lineage (3.1–4.5% nt and 3.7–3.9% aa) diversity (Table S3).

Alignment of the aa sequences of all viruses belonging to the six lineages of GIII revealed three non conservative changes, S124P (domain II), A169T (domain I) and K383 N (domain III). The 124S in lineage F and E was substituted by P in one isolate of lineage E (L11424_India_1984) and all the newer lineages (A–D of GIII representing 1989–2010). Furthermore, 124S/L was present in isolates of GI (Indonesia/Pacific) and GIV (Puerto Rico). The substitution, 124P was present in GII (South-east Asia). In domain I, 169A, present in lineage F was substituted by T in lineages A–E of GIII. 169A/V was present in GI, GII and GIV. In domain III, K383 present in lineage F was substituted by N in lineages A–E of GIII. 383 K was present in all other genotypes.

Two short regions with aa residues unique to the human isolate (NIV_664481_1966) of lineage F were observed: 210–243 in domain II with eight changes, two non-conservative (L215T and P226S) and 391–404 in transmembrane region with two changes, one non-conservative (A404S). Similarly the stretch of 150–196 residues (domain I & II) with five changes, four non conservative (L150Q G175E, T194M, and S196L) was unique to the mosquito isolate (NIV_664482_1966) of lineage F.

DENV-4: The E gene sequences of the fourteen Indian isolates (eleven sequenced in the present study and three retrieved from GenBank) were analyzed along with those of 40 publicly available global isolates. The Indian isolates were distributed into two genotypes, GI and a new genotype, GV. The recent isolates clustered into GI, lineage A (Fig. 2). The old Indian isolates from the 1960's along with one Thailand isolate from 1963 clustered independently, and showed high diversity from the other DENV-4 genotypes (Table S4). Considering 6% divergence as the minimum difference between genotypes, the cluster formed a new genotype, GV. DENV-4 therefore has five genotypes including sylvatic as GIV and GI with three lineages (A–C).

Analysis of nt diversity revealed that intra-genotypic divergence ranged from 0.3 to 3.1% and aa divergence ranged from 0.2 to 3.6% for GII to GV. GI with three lineages was found to be most divergent with 5.4% nt and 14.8% aa diversity. Inter-genotypic divergence for GI, II, III and V ranged from 6.8 to 10.5% in nt and 3.5 to 10% in aa sequences (Table S4). The sylvatic genotype, GIV was most diverse; it showed a nt diversity of 15.6–17.4% and aa diversity of 4.1–10.6% from the other genotypes. Most of the nt changes (>80%) were transitions and did not result in codon change. Comparison of the consensus aa sequences for each of

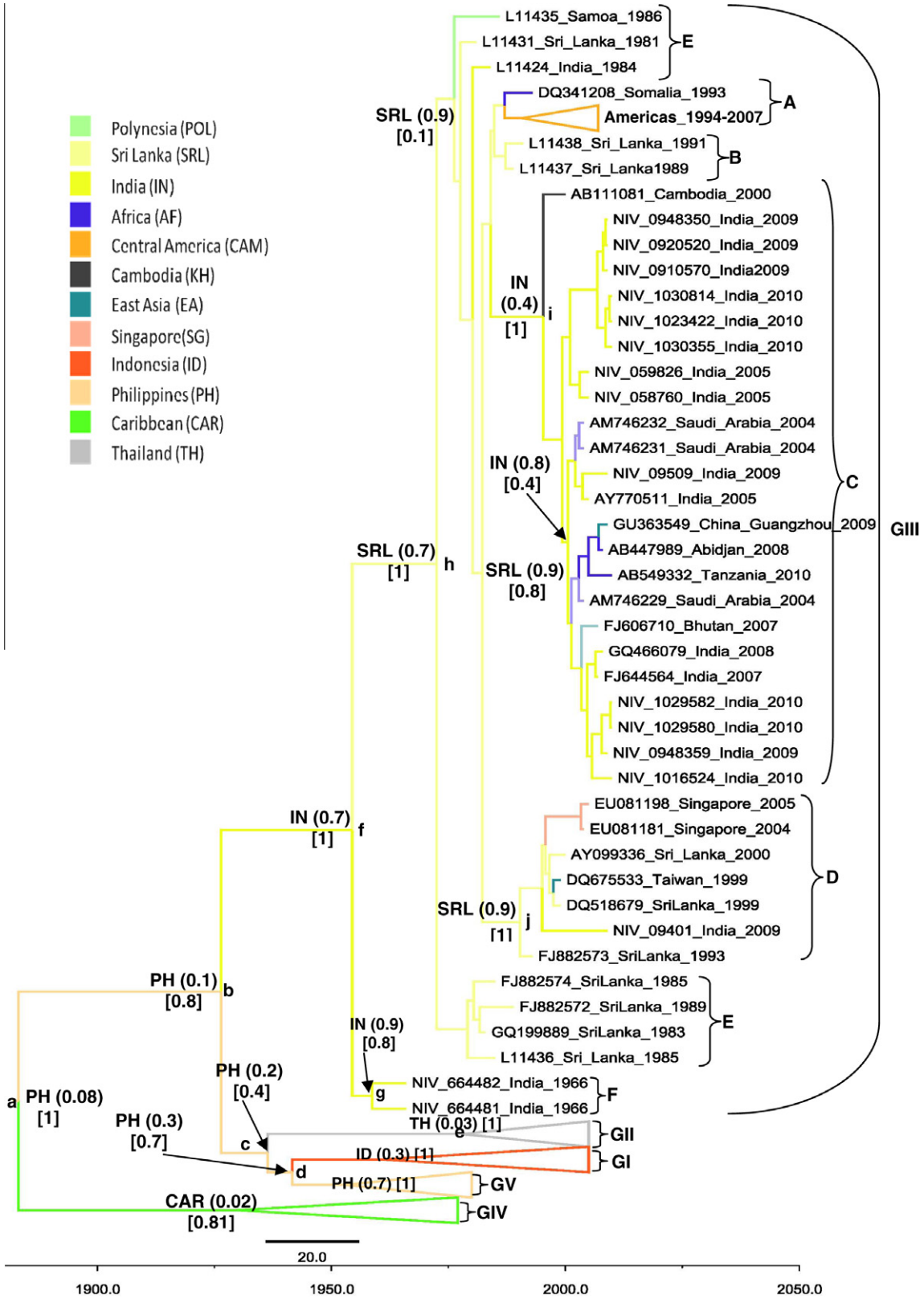

Fig. 1. Maximum clade credibility (MCC) tree of E gene sequences of DENV-3 with relaxed uncorrelated exponential clock with constant population size. The names of DENV-3 isolates include GenBank accession number, reference to country origin and year of sampling. The branches are colored according to the respective geographical region. Ancestral states with their probabilities are shown at key nodes (labeled a–j). Posterior supports are indicated in square brackets. The scale bar indicates time in years.

the genotypes with that of the sylvatic genotype revealed 35 substitutions with the largest number in domain III, residues 310–384 (Table S5). Sixteen changes were conserved across the genotypes. Seven unique with 4 non conservative changes (K174E, K202E, Y233H and S474P) were observed for GII which consisted of recent isolates from South Pacific region indicating it to be the most

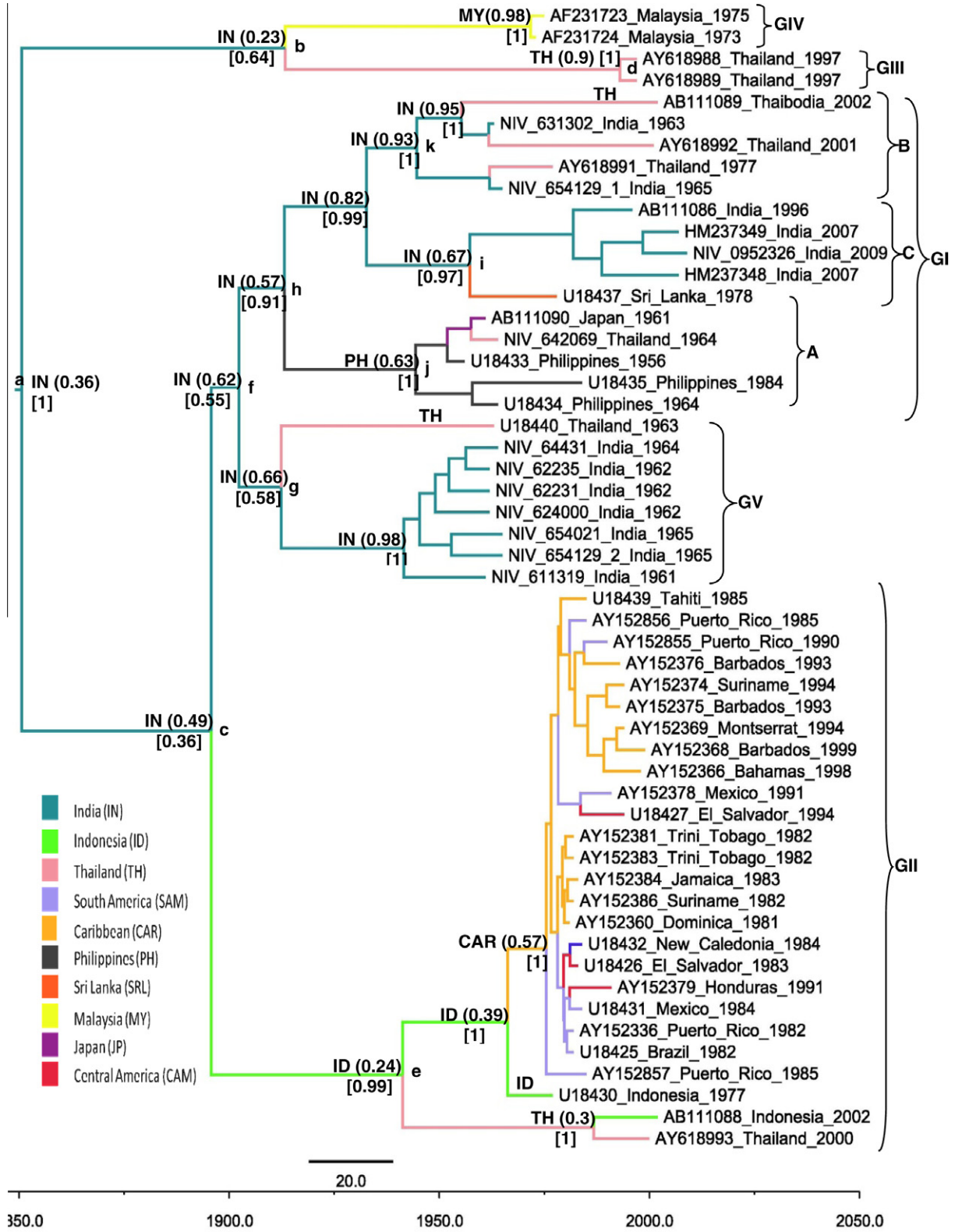

Fig. 2. Maximum clade credibility (MCC) tree of E gene sequences of DENV-4 with relaxed uncorrelated exponential clock with constant population size. The names of DENV-3 isolates include GenBank accession number, reference to country origin and year of sampling. The branches are colored according to the respective geographical region. Ancestral states with their probabilities are shown at key nodes (labeled a–k). Posterior supports are indicated in square brackets. The scale bar indicates time in years.

evolved genotype from GIV. GV the extinct Indian genotype harbored five unique substitutions and GI the current Indian genotype had 4, one of which was shared with GV. The changes in the Indian isolates were conservative in nature. The second putative glycosylation site at aa position 153 was absent in two Indian viruses, NIV_62235_1962 and NIV_64431_1964, belonging to GV.

3.2. Recombination analysis

Lineage/genotype switches observed by phylogenetic analysis suggested the possibility of a recombination event. All the E gene sequences, 113 of DENV-3 and 54 of DENV-4 were analyzed by GARD. No recombination event was detected for the DENV-3 isolates by GARD. However, SIMPLOT analysis suggested exchange of 4 short fragments between L11424_India_1984 and L11438_Sri Lanka_1991 resulting in the putative recombinant, NIV_09401_2009 (Fig. S1). For DENV-4, two breakpoints at position 777 and 1176 were suggested in the isolate, NIV_793679_India_1979. In the NJ tree constructed using three fragments, the best results were obtained with the 1–776 fragment (Fig. 3a). Using the 1–777 fragment of E, NIV_793679_1979 clustered into GV (98% bootstrap support) (Fig. 3b1) and shifted to GI when the 776–1500 fragment was used (Fig. 3b2). Simplot analysis revealed NIV_611319_1961 (GV) as one parent and Sri Lanka 1978 (U18437) (GI) as the other parent. The recombinant sequence was excluded from molecular clock analysis.

3.3. Molecular clock and phylogeography

Estimates of the nt substitution rates and the time to most recent common ancestors (tMRCA) for DENV-3 were inferred from 113 and for DENV-4 from 53 dated E gene sequences. Under the relaxed uncorrelated exponential clock with constant growth population model (Table S6), the mean substitution rate was 9.0×10^{-4} substitution/site/year(S/S/Y) [95% highest probability density (HPD) limits: 6.9×10^{-4}–1.0×10^{-3}]. For DENV-4, the mean substitution rate was 6.9×10^{-4} S/S/Y [95% HPD limits: 4.1×10^{-4}–1×10^{-3}]. The maximum clade credibility tree generated under the best fit model is shown in Figs. 1 and 2.

DENV-3: DENV-3 is the only serotype for which no sylvatic isolates are available. The estimate of the tMRCA for all genotypes of DENV-3 was about 127 years ago (node *a*) (95% HPD: 74, 199 years) with respect to the Indian isolates of 2010, correlating to the period from 1811–1936. GIV representing non-circulating viruses of the Caribbean region were the oldest and formed an out-group. All other genotypes, shared a common ancestor at node *b* (95% HPD: 1916–1953). GII, GI and GV diverged from a common node *c* (95% HPD: 1896–1949). The time for the divergence of GI/GV was 1925–1953 (node *d*), that of GII was during 1973–1981 (node *e)* and GIII was 1942–1963 (node *f*). Reconstruction of the ancestral states by phylogeographic analysis could not resolve the common ancestral state of all DENV-3 genotypes as the state probability was very low (Fig. 1). The same was not estimated by Araújo et al. (2009) as they investigated the three main genotypes separately. The most parsimonious ancestral state was determined to be Indonesia for GI, Thailand for GII, Caribbean (CAR) for GIV and Philippines for GV.

An evolutionary pattern of branching was observed in GIII. The ancestral state of GIII with lineages A–F was predicted to be India with a sp value of 0.7 (node *f*). A branch that emerged terminated in lineage F with India as its ancestral state (node *g*, sp = 0.8) and tMRCA 51 yrs ago (95% HPD: 1946–1958*)*. The other branch developed into all further lineages from node *h* with Sri Lanka as the ancestral state (sp = 0.9). There was a sequential emergence of lineages E, D, A and B with Sri Lanka as the ancestral state (sp = >0.7 at different nodes). Lineage E was not a single cluster, instead it had four branches and the time for its emergence was estimated to be 1964–1979 (node *h*). The tMRCA of lineage A consisting of all American isolates dated to about 1994–2007. The tMRCA of lineage D dated to 1987–1993 (node *j*). The tMRCA of lineage C, wherein majority of the Indian isolates clustered, dated to the period 1991–1999 (node *i*). India was the suggested ancestral state for lineage C (sp = 0.4) and for the other sub-clusters within lineage C (sp = 0.7–0.9). The data from India is fragmentary with no representation from 1966 to 2005 except for the single 1984 isolate (Fig 1). The transition of the isolates was defined by SPREAD and the transitions ($n = 36$) with BF values (>12) are presented in Table S7A and illustrated in Fig. 4 and Fig. S2 which shows that the virus has moved between continents. The highest BF values were for the transition between Indian/Middle-East (BF = 2838), followed by South America/Caribbean (BF = 547) and India/Sri Lanka (BF = 237).

DENV-4: The estimated tMRCA for all genotypes of DENV-4 was about 156 years with respect to the Indian isolates of 2009 (node *a*, 95% HPD: 1719–1931) (Fig. 2). India was predicted to be the ancestral state for the entire DENV-4 serotype albeit at low probability (sp = 0.36). India was the suggested ancestral state for GIII and GIV (sylvatic) genotypes (node *b* sp = 0.23) with tMRCA during 1856–1953 and for GI/GV and GII genotypes (node *c*, sp = 0.49,

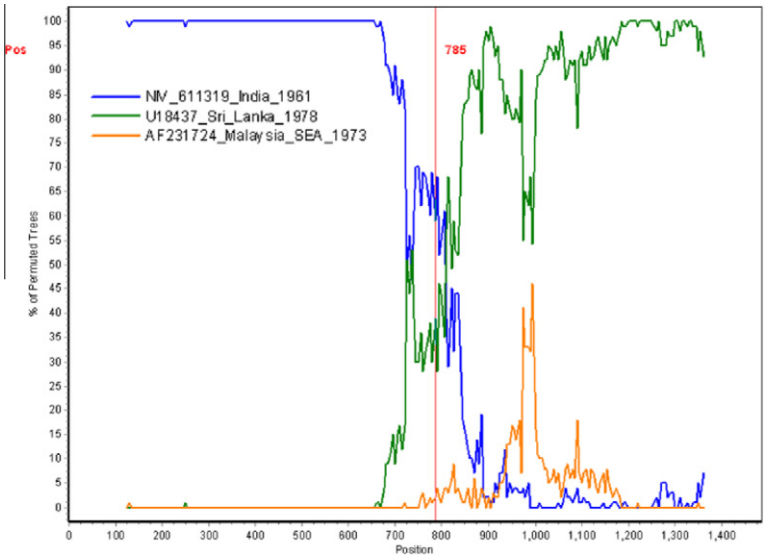

Fig. 3a. Bootscanning analysis to confirm the putative recombinant event. The X-axis indicates the nucleotide positions of the viral genome. The Y-axis represents the bootstrap support for the grouping of the NIV_793679_1979 recombinant strain with either parental strain.

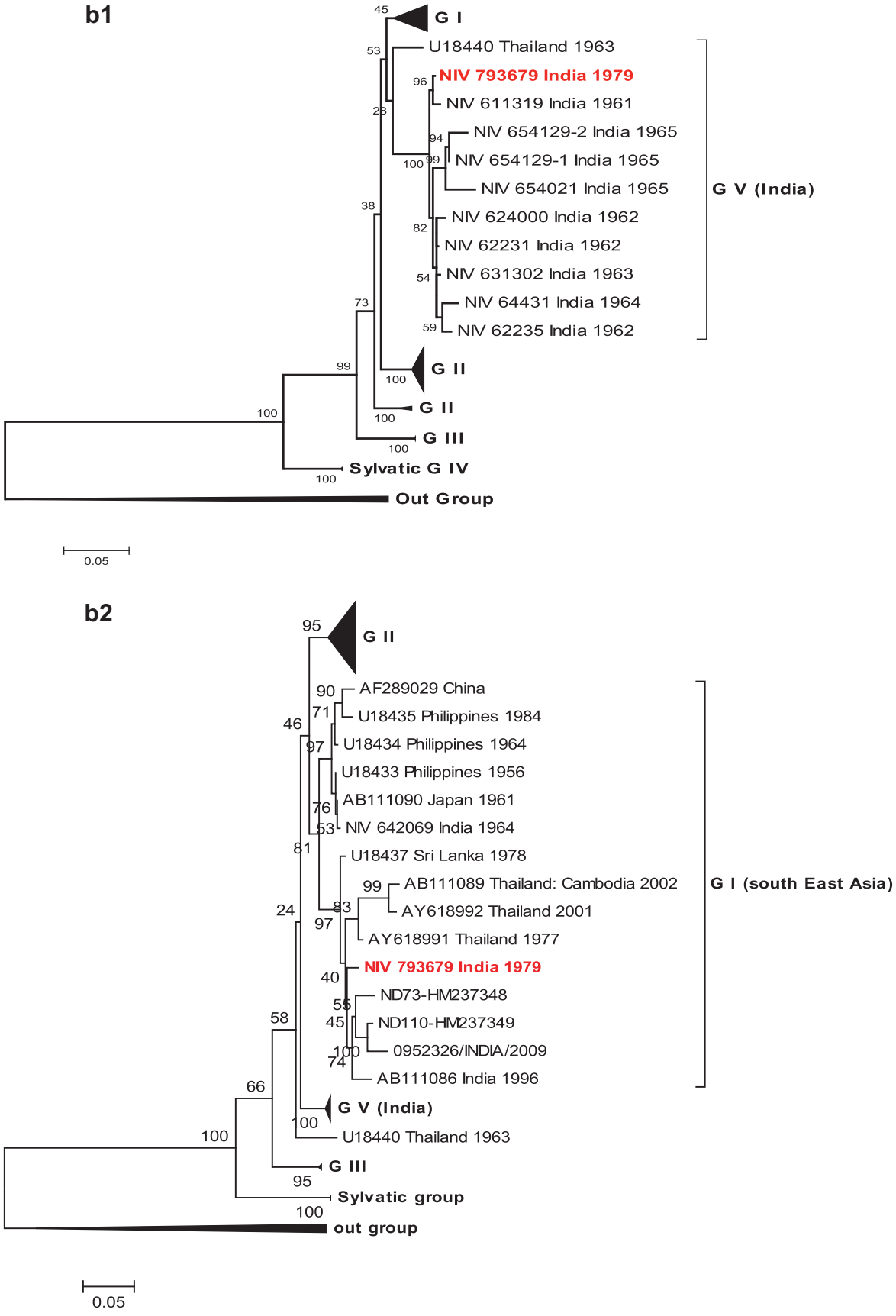

Fig. 3b. (1 and 2). NJ trees showing the changing phylogenetic position of NIV_793679_India_1979 from genotype V to I at breakpoint 786.

tMRCA 1839–1932). The tMRCA of the Thai isolates of GIII dated to 1997–2007 (node *d*) and GIV dated to 1969–1973. Further, Indonesia was suggested as the ancestral state of GII containing viruses from the Pacific region (node *e*, sp = 0.24, tMRCA of

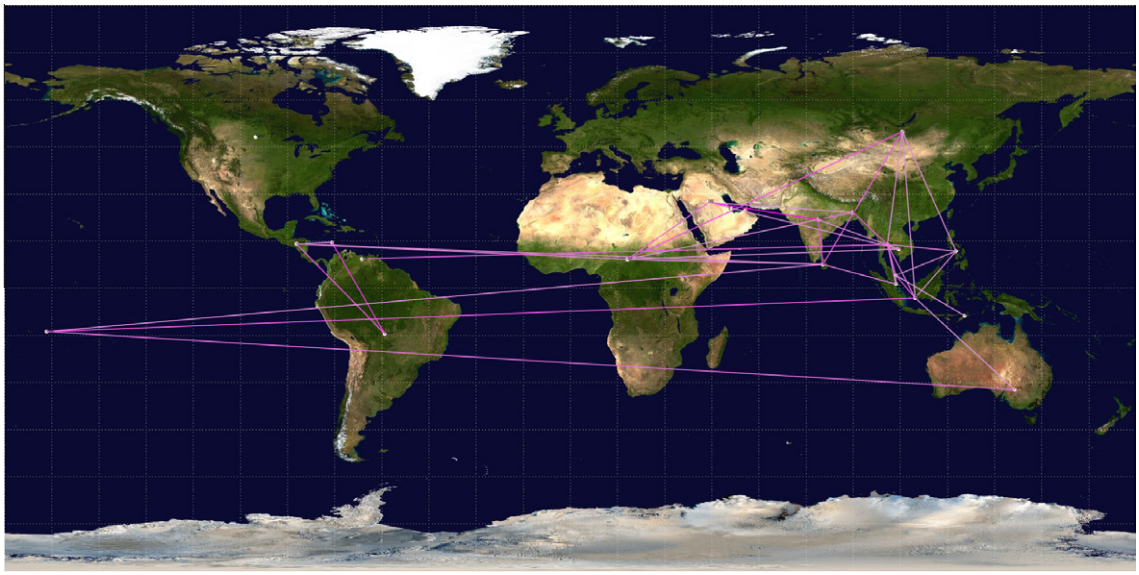

Fig. 4. Map showing significant migration links between different geo-regions for DENV-3 as supported by BF value >12. Darker shades of pink correspond to transition rates having higher BF values.

1901–1970) and India was at the ancestral node of GI and GV which contained Indian isolates (node *f*, sp = 0.62, tMRCA1847–1935). GV (node *g*, tMRCA 1863–1944) was populated exclusively with Indian isolates from 1960s and one isolate from Thailand (1963). The absence of recent isolates suggests that the GV is an extinct genotype. GI (tMRCA 1869–1944, node *h*) branched into the 3 lineages - A, B and C, which had tMRCAs dating to 1929–1956, 1931–1957, and 1936–1975, respectively. Lineage A with Philippines as the ancestral state (sp = 0.63) represented the Far East. Lineage B and Lineage C with India as the ancestor (sp = 0.93 and 0.97, respectively) represented Thai, Indian, Sri Lankan and Cambodian isolates.

The transition of the isolates with BF values (>3) are presented in Table S7B and illustrated in Fig. 5 and Fig. S3. Fewer transitions ($n = 15$) were revealed and the BF values were much lower than that observed for DENV-3. The highest BF values were for the transition between South America/Caribbean (BF = 571) followed by India/Thailand (BF = 19.4) and South America/Central America (BF = 11.4).

3.4. Selection pressure analysis

Sites were considered to be under positive selection if proved by at least two of the methods with high statistical significance ($P < 0.1$/ Bayes factor >50). No selection pressure was noted within the Indian isolates of DENV-3 and 4 or when individual genotypes were analyzed (data not shown). Weak evidence of positive selection was observed only by the FEL method for GIII isolates of DENV-3 at residues 122, 221 (domain II), 178 (domain I) and for GI/GV isolates of DENV-4 at residues 132, 270 (domain II), 169, (domain I) and 380 (domain III) in a combined analysis.

4. Discussion

Our present study on DENV-3 and 4 combined with our earlier studies on DENV-1 and 2 (Kumar et al., 2010; Patil et al., 2011) prove that the population of dengue viruses presently circulating in India is different from that in the 1960s. The Indian isolates of DENV-3 from the 1960s and the present times belonged to GIII, which has reportedly five lineages, A–E (Sharma et al., 2011). Two isolates, one mosquito and one human from South India, representing the 1960s, formed a new lineage, F. The high sequence diversity observed within lineage F could be attributed to the host difference whilst the diversity between lineages implied its evolutionary distance from the currently circulating viruses. Lineage E along with lineage F represented the time period when the disease was mild. Lineages B, C and D represented the time period when disease presentation was more severe. Lineage C had the broadest distribution including the Indian subcontinent, the Middle East and Africa. Lineage A had a narrow range of geographical distribution with isolates from the Americas.

Three non-conservative aa changes, S124P (domain II), A169T (domain I) and K383 N (domain III), were detected in lineages A–D of GIII. Whilst 124P was shared with GII, 169T and 383 N were specific to lineages A–D of GIII. Considering that GII and GIII (lineages A–D) represent viruses currently circulating in most of the countries and associated with severe disease, the S124P mutation may be responsible for making the virus more fit and perhaps virulent. An interesting observation was the presence of one isolate of 1984 from India in lineage E which had the S124P substitution without the two other changes and therefore could represent the transition phase from the old to the new. Substitution at 124 was also reported to be associated with DHF in Indonesia by Dewi et al. (2009).

In addition, there were aa residues unique to the isolates of the extinct Lineage F. The stretch of aa residues, 210–243 had 8 residues and 391–404 had 2 residues unique to the human isolate and the stretch from aa 150–196 had five residues unique to the mosquito isolate. The 210–243 stretch is a part of predicted consensus Th-cell epitopes (Mazumder et al., 2007) and 391–404 stretch is localized on the upper lateral surface of domain III and consists of heparin sulfate binding domains residues (Chen et al., 1997). The stretch 150–196 is a part of domain II which is the region of dimerization and important to the fusion event (Heinz and Stiasny, 2006). The results imply that mutations that occurred at these sites conferred fitness to the viruses that subsequently evolved.

The molecular clock analysis revealed the tMRCA to be ~127 years old (~1883) for DENV-3 with an estimated evolutionary rate of 9.0×10^{-4} S/S/Y, similar to that reported by Araujo et al. (2009). DENV-3, at the oldest node, diversified into two main branches one that seeded Pacific region (GIV) and became extinct and the second diverged to populate South Asia/South Pacific (GI/GV), Thailand (GII) and the Indian subcontinent (GIII).

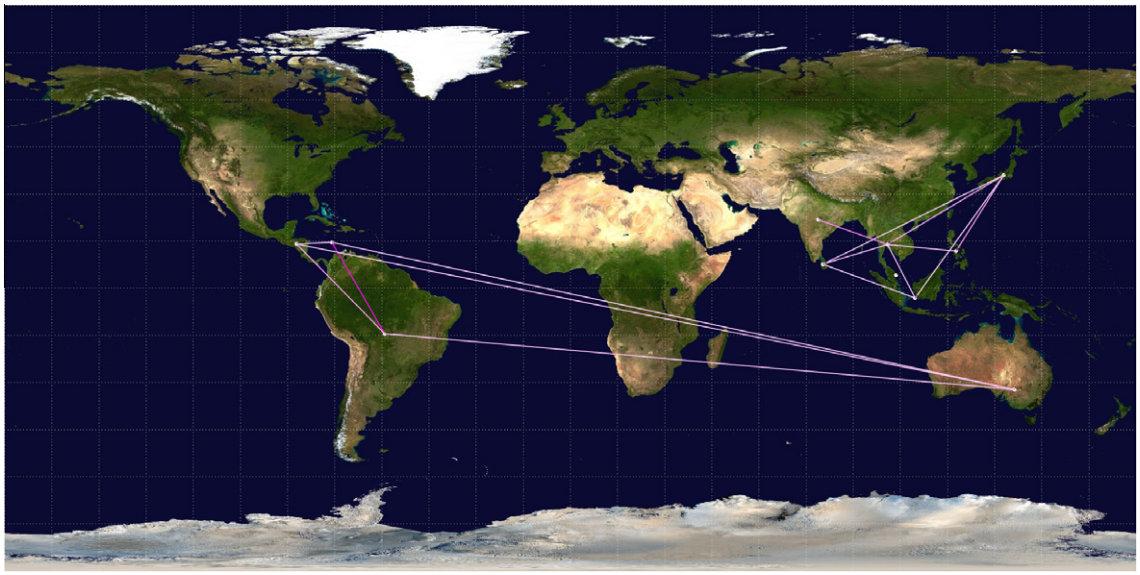

Fig. 5. Map showing significant migration links between different geo-regions for DENV-4 as supported by BF value >3. Darker shades of pink correspond to transition rates having higher BF values.

The ancestral state of GIII (node *f*) was predicted to be India during 1942–1963, which differs from the earlier prediction of Sri Lanka (1967–1979) by Araujo et al. (2009). Two main branches emerged from node *f*, one terminated in India (lineage F) and the other continued in Sri Lanka, lineage E onwards. During 1981–88 an introduction into India from Sri Lanka which evolved into lineage C from node *i* during 1991–1999 was suggested. Therefore, two events, one exportation from India (pre-DHF period) and one importation from Sri Lanka (post-DHF period) occurred. Movement of DENV-3 between the two countries was supported by a high BF value. It is possible that additional events have been masked by the paucity of data from India during 1970–2000.

During the latter period (1999–2003) the virus was likely to have been exported from India to Saudi Arabia, Africa and China. BF analysis provided significant evidence for movement of the virus between India and the Middle East. It has been suggested that pilgrimage to Haj which was subsidized by the Indian government since 1993, might be instrumental in the importation of viruses to the Middle East (Zaki et al., 2008). Investigations in Bhutan during the 2006–2007 outbreaks indicated introduction of lineage C from northern India (Dorji et al., 2009). Broadcasting of the virus from India to Cambodia was also indicated with a significant BF value. On the other hand, Bangladesh and Myanmar were populated with GII viruses from South-east Asia as suggested earlier (Podder et al., 2006) and also shown from the present phylogeography studies. Further, dispersal of the virus from Sri Lanka to Africa and Central America was also supported by significant values of BF (Table S7A).

Furthermore, the role of lineage E in the origination of lineages A, B, C and D was suggested by the characteristics of the L11424_India_1984 isolate. The virus harbored the S124P mutation seen in lineages A–D and was suggested to be a putative parent of NIV_09410_2009 by SIMPLOT analysis.

For DENV-4 the data revealed that the genotype changed from GV (1960s) to GI (1979 onwards) in India. GI was recently detected in Brazil for the first time (Melo et al., 2009), a region that was earlier populated with GII. The genomic diversity observed within GI was highest and suggests that it may be an evolving lineage. In contrast, GII which was most diverse from the sylvatic GIV indicates that it is an evolved lineage. The predicted N linked glycosylation site at position 67 was conserved among all DENV-4 viruses, however, the second putative glycosylation site at aa position 153 was absent in two Indian viruses of 1960s belonging to GV. Similar findings were reported by Lanciotti et al. (1997) for DENV-4 isolates from Thailand (1963), Indonesia (1973) and Dominica (1981).

Recombination analysis revealed that the Indian 1979 isolate of GI was a recombinant of the Indian 1961 isolate of GV and the Sri Lankan 1978 isolate of GI. Inter-genotypic recombination of GI and GII was earlier reported by AbuBakar et al. (2002) and Worobey et al. (1999) for DENV-4.

GI was divided into three lineages (Cecilia et al., 2011). All the Indian isolates, from 1979 to 2009 clustered within lineage A along with Sri Lankan isolates, while the Thai and Far East isolates formed separate lineages B and C respectively.

The molecular clock analysis of 53 DENV-4 isolates, excluding the recombinant virus revealed that the rate of nucleotide substitution (6.9×10^{-4} S/S/Y) as well as tMRCA for DENV-4 ($\sim$156 years, $\sim$1853) matched with earlier reports (Twiddy et al., 2003).

Reports so far had placed GIV (sylvatic) at the root of DENV-4 (Lanciotti et al., 1997; Dash et al., 2011) based on phylogenetic analysis. The addition of data on Indian isolates from the 1960s and the use of phylogeography analysis placed India at the root of DENV-4 viruses during 1719–1931, a period that coincides with movement of people between Africa and Asia. Phylogeography by Villabano and Zanotta, (2011), which did not have the data from India, showed Malaysia/Thailand as the ancestor state. GIII, an extinct branch representing Thailand, diverged around 1997–2007. GV represented the extinct cluster of Indian viruses of the 1960s which shared a common ancestor (India, 1863–1944) with a Thailand isolate, implying movement of viruses from India to Thailand. Bayes factor analyses of the transition rates (Fig. 4) also supported the virus movement between India and Thailand. GI represented currently circulating viruses from India and South East Asia. India was predicted to be the ancestral state at almost all nodes within GI except node *j*, indicating that viruses from India possibly spread to Thailand, Sri Lanka, Cambodia and Philippines. Philippines then probably exported the virus to Japan and Thailand. BF analysis supported both these transitions. GII branched out from node *c* (India 1839–1932) and was rooted at node *e* (1901–1970) represented by Indonesia which then possibly populated Thailand, supported by a significant BF value. Indonesia may also have spread the virus to the Americas via the Caribbean. A significant Bayes value was noted for the Caribbean and Central America transition. Exchanges

are also indicated between Central America, Australia and South America (Table S7B).

Overall, the transportation of viruses between Sri Lanka and India was evident for both DENV-3 and DENV-4 during the 1970s and 1980s which correlated to human movements during the Sri Lankan civil war. It was suggested by Twiddy et al. (2003) that dengue virus separated into distinct serotypes independently because of geographical partitioning of different primate populations. The present study introduces the interesting concept that DENV-4 may have originated in India. The origin of DENV-3 could not be ascertained from the study and additional data on sylvatic strains are required.

5. Conclusion

Our earlier studies on the molecular evolution of DENV-2 (Kumar et al., 2010) and DENV-1 (Patil et al., 2011) and the present study on DENV-3 and 4 in summary prove that the population of dengue viruses in India has undergone major changes since the 1960s with either genotype or lineage changes. Fresh introductions, evolution in situ along with limited recombination have contributed to diversity and the dengue evolutionary dynamics in the country. The present study revealed the major role of India in the dispersion of DENV-3 and 4. Also important was the probable role of exchange of viruses between India and Sri Lanka in the evolution of virulent strains of DENV-3.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.meegid.2012.07.009. These data include Google maps of the most important areas described in this article.

References

AbuBakar, S., Wong, P.F., Chan, Y.F., 2002. Emergence of dengue virus type 4 genotype IIA in Malaysia. J. Gen. Virol. 83 (10), 2437–2442.

Araújo, J.M., Nogueira, R.M., Schatzmayr, H.G., Zanotto, P.M., Bello, G., 2009. Phylogeography and evolutionary history of dengue virus type 3. Infect. Genet. Evol. 9 (4), 716–725.

Bielejec F., Rambaut A., Suchard M.A, Lemey P. SPREAD: Spatial Phylogenetic Reconstruction of Evolutionary Dynamics. *Bioinformatics*, 2011.

Carey, D.E., Myers, R.M., Reuben, R., 1964. Dengue Types 1 and 4 Viruses in Wild-Caught Mosquitoes in South India. Science 143 (3602), 131–132.

Cecilia, D., Kakade, M.B., Bhagat, A.B., Vallentyne, J., Singh, A., Patil, J.A., Todkar, S.M., Varghese, S.B., Shah, P.S., 2011. Detection of Dengue-4 virus in Pune, Western India after an absence of 30 years–its association with two severe cases. J. Virol. 8 (1), 46.

Chen, Y., Maguire, T., Hileman, R.E., Fromm, J.R., Esko, J.D., Linhardt, R.J., Marks, R.M., 1997. Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. Nat. Med. 3 (8), 866–871.

Dar, L., Gupta, E., Narang, P., Broor, S., 2006. Cocirculation of dengue serotypes, Delhi, India, 2003. Emerg. Infect. Dis. 12 (2), 352–353.

Dash, P.K., Parida, M.M., Saxena, P., Kumar, M., Rai, A., Pasha, S.T., Jana, A.M., 2004. Emergence and continued circulation of dengue-2 (genotype IV) virus strains in northern India. J. Med. Virol. 74 (2), 314–322.

Dash, P.K., Parida, M.M., Saxena, P., Abhyankar, A., Singh, C.P., Tewari, K.N., Jana, A.M., Sekhar, K., Rao, P.V., 2006. Reemergence of dengue virus type-3 (subtype-III) in India: implications for increased incidence of DHF & DSS. J. Virol. 6 (3), 55.

Dash, P.K., Sharma, S., Srivastava, A., Santhosh, S.R., Parida, M.M., Neeraja, M., Subbalaxmi, M.V., Lakshmi, V., Rao, P.V., 2011. Emergence of dengue virus type 4 (genotype I) in India. Epidemiol. Infect. 139 (6), 857–861.

Dewi, B, E., Tomohiko, T., Shigeru, T., T Mirawati, S., R.P. Larasati c., Andrew, L., C., and Ichiro K., 2009. Genotypic and phenotypic characteristics of DENV-3 isolated from patients with different disease severities in Indonesia. Dengue Bulletin – Vol. 33.

Dorji, T., Yoon, I.K., Holmes, E.C., Wangchuk, S., Tobgay, T., Nisalak, A., Chinnawirotpisan, P., Sangkachantaranon, K., Gibbons, R.V., Jarman, R.G., 2009. Diversity and origin of dengue virus serotypes 1, 2, and 3 Bhutan. Emerg. Infect. Dis. 15 (10), 1630–1632.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.

Gupta, E., Dar, L., Kapoor, G., Broor, S., 2006. The changing epidemiology of dengue in Delhi India. J. Virol. 5 (3), 92.

Hang, T.V., Holmes, E.C., Veasna, D., Quy, N.T., Tinh, H.T., 2010. Emergence of the asian 1 genotype of dengue virus serotype 2 in Viet Nam: *in vivo* fitness advantage and lineage replacement in South-East Asia. PLoS Negl. Trop. Dis. 4 (7), e757.

Heinz, F.X., Stiasny, K., 2006. Flavivirus membrane fusion. J. Gen. Virol. 87, 2755–2766.

Kimura, R., Hotta, S., 1943. Experimental inoculation of dengue virus into mice. Nippon Igaku. 3344, 1378–1379.

Kosakovsky-Pond, S.L., Frost, S.D.W., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21, 2531–2533.

Kukreti, H., Dash, P.K., Parida, M., Chaudhary, A., Saxena, P., Rautela, R.S., Mittal, V., Chhabra, M., Bhattacharya, D., Lal, S., Rao, P.V., Rai, A., 2009. Phylogenetic studies reveal existence of multiple lineages of a single genotype of DENV-1 (genotype III) in India during 1956–2007. J. Virol. 6 (6), 1.

Kumar, S.R., Patil, J.A., Cecilia, D., Cherian, S.S., Barde, P.V., Walimbe, A.M., Yadav, P.D., Yergolkar, P.N., Shah, P.S., Padbidri, V.S., Mishra, A.C., Mourya, D.T., 2010. Evolution, dispersal and replacement of American genotype dengue type 2 viruses in India (1956–2005): selection pressure and molecular clock analyses. J. Gen. Virol. 91 (3), 707–720.

Lanciotti, R.S., Gubler, D.J., Trent, D.W., 1997. Molecular evolution and phylogeny of dengue-4 viruses. J. Gen. Virol. 78 (9), 2279–2284.

Lemey, P., Rambaut, A, Drummond, A., J., Suchard, M., A., 2009. Bayesian phylogeography finds its roots. *PLoS Comput Biol.* Sep; 5(9):e1000520. Epub 2009 Sep 25.

Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray, S.C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. 73, 152–160.

Mazumder, R., Hu, Z.Z., Vinayaka, C.R., Sagripanti, J.L., Frost, S.D., Kosakovsky-Pond, S.L., Wu., C.H., 2007. Computational analysis and identification of amino acid sites in dengue E proteins relevant to development of diagnostics and vaccines. Virus Genes 35 (2), 175–186.

Melo, F.L., Camila, M.R., Paolo, M., Zanotto, A., 2009. Introduction of dengue virus 4 (DENV-4) genotype I into brazil from asia? PLoS Negl. Trop. Dis. 3 (4), e390.

Messer, W.B., Vitarana, U.T., Sivananthan, K., Elvtigala, J., Preethimala, L.D., Ramesh, R., Withana, N., Gubler, D.J., De Silva, A.M., 2002. Epidemiology of dengue in Sri Lanka before and after the emergence of epidemic dengue hemorrhagic fever. Am. J. Trop. Med. Hyg. 66 (6), 765–773.

Messer, W.B., Gubler, D.J., Harris, E., Sivananthan, K., de Silva, A.M., 2003. Emergence and global spread of a dengue serotype 3, subtype III virus. Emerg. Infect. Dis. 9 (7), 800–809.

Meyers, R.M., Varkey, M.J., Reuben, R., Jesudas, E.S., 1970. Dengue outbreak in Vellore, Sourthern India, in 1968, with isolation of four dengue types from man and mosquitoes. Indian J. Med. Res. 58, 168–173.

Pan American Health Organization, Regional Program on Dengue through the PAHO/WHO Representation in Paraguay (PAHO-Paraguay). Ministry of Public Health and SocialWelfare (Ministerio de Salud Publica y Bienestar Social/MSPBS) of Paraguay in EID Updates, Region of the Americas, 2007. Dengue Outbreak Continues to Subside in Paraguay. Emerging and Reemerging Infectious Diseases 4

Passos, M., Santos, L., Pereira, M., Casali, C.G., Fortes, B., Ortiz Valencia, L., Alexandre, J., Medronho, R., 2004. Clinical differences observed in patients with dengue caused by different serotypes in the epidemic of 2001/2002, occurred in Rio de Janeiro. Rev. Soc. Bras. Med. Trop. 37, 293–295.

Patil, J.A., Cherian, S., Walimbe, A.M., Patil, B.R., Sathe, P.S., Shah, P.S., Cecilia, D., 2011. Evolutionary dynamics of the American African genotype of dengue type 1 virus in India (1962–2005). Infect. Genet. Evol. 11 (6), 1443–1448.

Peláez, O., Guzmán, M.G., Kouri, G., Pérez, R., San Martín, J.L., Vásquez, S., Rosario, D., Mora, R., Quintana, I., Bisset, J., Cancio, R., Mesa, A.M., Castro, O., González, D., Avila, L.C., Rodriguez, R., Alvarez, M., Pelegrino, J.L., Bernardo, L., Prado, I., 2004. Dengue 3 Epidemic, Havana, 2001. Emerg. Infect. Dis. 10, 719–722.

Podder, G., Breiman, R.F., Azim, T., Thu, H.M., Velathanthiri, N., Mai le, Q., Lowry, K., Aaskov, J.G., 2006. Origin of dengue type 3 viruses associated with the dengue outbreak in Dhaka, Bangladesh, in 2000 and 2001. Am. J. Trop. Med. Hyg. 74 (2), 263–265.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Pushpa, V., Venkatadesikalu, M., Mohan, S., Cherian, T., John, T.J., Ponnuraj, E.M., 1998. An epidemic of dengue haemorrhagic fever/dengue shock syndrome in tropical India. Ann. Trop. Paediatr. 18 (4), 289–293.

Rahman, M., Rahman, K., Siddque, A.K., Shoma, S., Kamal, A.H., Ali, K.S., Nisaluk, A., Breiman, R.F., 2002. First outbreak of dengue hemorrhagic fever Bangladesh. Emerg. Infect. Dis. 8 (7), 738–740.

Rico-Hesse, R., 2010. Dengue virus virulence and transmission determinants. Curr. Top. Microbiol. Immunol. 338, 45–55, Review.

Sabin, A.B., 1952. Research on Dengue during World War II. Am. J. Trop. Med. Hyg. 1, 30–50.

Salminen, M.O., Carr, J.K., Burke, D.S., McCutchan, F.E., 1995. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. AIDS Res. Hum. Retroviruses 11, 1423–1425.

Sharma, S., Dash, P.K., Agarwal, S., Shukla, J., Parida, M.M., Rao, P.V., 2011. Comparative complete genome analysis of dengue virus type 3 circulating in India between 2003 and 2008. J. Gen. Virol. 92 (7), 1595–1600.

Suchard, M.A., Weiss, R.E., Sinsheimer, J.S., 2001. Bayesian selection of continuoustime Markov chain evolutionary models. Mol. Biol. Evol. 18, 1001–1013.

Sugiura, N., 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. Comm. Statist. 7, 13–26.

Twiddy, S.S., Holmes, E.C., Rambaut, A., 2003. Inferring the rate and time scale of dengue virus evolution. Mol. Biol. Evol. 20, 122–129.

Villabona-Arenas, C.J., Zanotto, P.M., 2011. Evolutionary history of Dengue virus type 4: insights into genotype phylodynamics. Infect. Genet. Evol. 11 (5), 878–885.

Worobey, M., Holmes, E.C., 1999. Evolutionary aspects of recombination in RNA viruses. J. Gen. Virol. 80, 2535–2543.

Worobey, M., Rambaut, A., Holmes, E.C., 1999. Widespread intra-serotype recombination in natural populations of dengue virus. Proc. Nat. Acad. Sci. USA 96, 7352–7357.

Zaki, A., Perera, D., Jahan, S.S., Cardosa, M.J., 2008. Phylogeny of dengue viruses circulating in Jeddah, Saudi Arabia: 1994 to 2006. Trop. Med. Int. Health 13, 584–592.

Zhang, C., Mammen, M.P., Chinnawirotpisan, P., Klungthong, C., Rodpradit, P., Monkongdee, P., Nimmannitya, S., Kalayanarooj, S., Holmes, E.C., 2005. Prevalence Virol. 79 (24), 15123–15130.