

Evolution, dispersal and replacement of American genotype dengue type 2 viruses in India (1956–2005): selection pressure and molecular clock analyses

Sandeep R. P. Kumar,^{1†} Jayashri A. Patil,^{1†} D. Cecilia,¹ Sarah S. Cherian,¹ Pradip V. Barde,¹ Atul M. Walimbe,¹ Pragya D. Yadav,¹ Prasanna N. Yergolkar,¹ Paresh S. Shah,¹ Vasant S. Padbidri,² Akhilesh C. Mishra¹ and Devendra T. Mourya¹

Correspondence

Devendra T. Mourya
mouryadt@icmr.org.in

¹National Institute of Virology, 130/1 Sus Road, Pashan, Pune – 411021, Maharashtra, India

²King Edward Memorial Hospital and Research Center, 489, Rasta Peth, Sardar Moodliar Road, Pune – 411011, Maharashtra, India

This study reports the phylogeny, selection pressure, genotype replacement and molecular clock analyses of many previously unstudied dengue type 2 virus (DENV-2) strains, isolated in India over a time span of almost 50 years (1956–2005). Analysis of complete envelope (E) gene sequences of 37 strains of DENV-2 from India, together with globally representative strains, revealed that the American genotype, which circulated predominantly in India during the pre-1971 period, was then replaced by the Cosmopolitan genotype. Two previously unreported amino acid residues, one in the American (402I) and one in the Cosmopolitan (126K) genotypes, known to be involved functionally in the cellular tropism of the virus, were shown to be under positive selection pressure. The rate of nucleotide substitution estimated for DENV-2 was 6.5×10^{-4} substitutions per site year⁻¹, which is comparable with earlier estimates. The time to the most recent common ancestor of the pre-1971 Indian strains and the American genotype was estimated to be between 73 and 100 years (1905–1932), which correlates with the historical record of traffic between India and South America and suggests transportation of the virus from the Americas. Post-1971 Indian isolates formed a separate subclade within the Cosmopolitan genotype. The estimated time to the most recent common ancestor of the Indian Cosmopolitan strains was about 47 years, with further estimates indicating the migration of DENV-2 from India to countries across the Indian ocean between 1955 and 1966. Overall, the present study increases our understanding of the events leading to the establishment and dispersal of the two genotypes in India.

Received 29 October 2009

Accepted 12 November 2009

INTRODUCTION

Dengue is the most prevalent arthropod-borne viral infection of humans in tropical and subtropical countries. About 100 million infections occur every year, of which the majority are either asymptomatic or self-limiting, acute, severe, febrile, 'flu-like illnesses often accompanied by rash.

However, nearly half a million cases of dengue haemorrhagic fever or dengue shock syndrome (DHF/DSS) are estimated to be hospitalized throughout tropical and subtropical regions (WHO, 2002). A high proportion of patients who develop DSS die. Therefore, dengue fever (DF) and/or DHF/DSS have become a major public-health concern in the tropics and subtropics worldwide. Dengue virus (DENV) infections are caused by one or more of four antigenically related DENV serotypes (1–4). DF is commonly described as being confined to urban areas. However, it is now commonly reported in rural settings in both Asia and Latin America (Eram *et al.*, 1979; Mehendale *et al.*, 1991; Mahadev *et al.*, 1993; Hayes *et al.*, 1996; Chareonsook *et al.*, 1999; Kumar *et al.*, 2001; Tewari

†These authors contributed equally to this work.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are FJ538905–FJ538928 and FJ807632–FJ807640.

Three supplementary tables are available with the online version of this paper.

et al., 2004; Arunachalam *et al.*, 2004). Although *Aedes aegypti* is the main vector for DENV, epidemiological evidence suggests that other *Aedes* species play an important role in DENV transmission (Mackerras, 1946; Rosen *et al.*, 1954; Sulianti, 1978; Gratz, 2004; Xu *et al.*, 2007; Fulmali *et al.*, 2008). Enzootic DENV may also be responsible for DF in rural and periurban settings throughout the tropics, except in the Americas (Diallo *et al.*, 2008).

DENV belongs to the family *Flaviviridae*, genus *Flavivirus*, and its genome consists of a non-segmented, single-stranded, positive-sense RNA, which is approximately 10.7 kb. It encodes a single open reading frame that encodes three structural (capsid, C; membrane, M; and envelope, E) and seven non-structural (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) proteins (Chambers *et al.*, 1990).

Although historical records state that DENV was present in India in the 19th century, no viruses were obtained until DENV-2 was first isolated from the sera of American soldiers in 1944 during World War II (Sabin, 1952). Since then, epidemics of DENV have been recorded from almost all over India at different times. Many factors – including unprecedented population growth, increased population density, unplanned and uncontrolled urbanization, increased global travel, increased density of the vector mosquito, infestation of new geographical areas by vector mosquitoes, warm and humid climate and water storage pattern in houses – that promote the spread of DENV have all contributed to the remarkably increased epidemiology of DENV in India during the past few decades (Chaturvedi, 2006). All four DENV serotypes have been reported periodically in this country, but DENV-2 has usually been the main aetiological agent in DF and DHF outbreaks and has emerged as the predominant serotype (Dar *et al.*, 1999; Parida *et al.*, 2002). Various studies have revealed genetic variation amongst DENV-2 isolates on the basis of E gene sequence analysis and grouped them into six genotypes corresponding to Sylvatic, Asian I, Asian II, American/Asian, Cosmopolitan and American, as classified previously (Rico-Hesse, 1990; Rico-Hesse *et al.*, 1997; Wang *et al.*, 2000). Some genotypes reported from multiple geographical regions demonstrate the wide distribution of these viruses, whilst others appear to have more restricted distributions. It has been proposed that DENV-2 genotypes differ in their virulence and in their ability to infect host cells and to be transmitted, thus influencing the outcome and possibly the dispersal of the disease caused by the virus (Leitmeyer *et al.*, 1999; Diamond *et al.*, 2000; Cologna & Rico-Hesse, 2003; Cologna *et al.*, 2005).

As there is little information regarding the circulating genotypes of DENV-2 in India and because vaccines are now realistically within reach, it is important to genotype and to understand the epidemiology of these viruses (Chaturvedi & Shrivastava, 2004). Limited phylogenetic studies of DENV-2, utilizing a small fragment of either the

E/NS1 gene or the capsid–premembrane (C–prM) gene, have been reported, showing a change in the circulating genotype of DENV-2 in India (Singh *et al.*, 1999; Singh & Seth, 2001; Dash *et al.*, 2004). These studies focused mainly on strains of DENV-2 from single epidemics. Studying the evolutionary dynamics of DENV in India over an extended period of time will be useful to understand how different genotypes were established in India and how they dispersed. As the genetic diversity following DENV-2 subtype replacement may have a significant role to play in the outcome of DENV infection and disease severity, we also looked for evidence of selection pressure and clade reintroduction in these viruses, based on previous observations with non-Indian DENV-2 (Zanotto *et al.*, 1996; Twiddy *et al.*, 2002a, b; Myat Thu *et al.*, 2005; Bennett *et al.*, 2006; Vasilakis *et al.*, 2007).

We report the molecular evolution, geographical dispersal and replacement of the American genotype based on 37 DENV-2 strains from India, sampled during the past 50 years, in the context of globally representative strains. The E gene was selected because it is known to be under immunological selection pressure imposed by the host immune system (Innis *et al.*, 1989; Gritsun *et al.*, 1995). Site-specific selection pressures and rates of nucleotide substitution were estimated.

RESULTS

Sequence and phylogenetic analyses

The percentage nucleotide identity (PNI) of the E gene of Indian strains, when compared with the prototype strain of DENV-2 (PG/NGC44/1944), ranged from 89.9 to 95.6%, and the deduced amino acid sequence identity ranged from 98.2 to 98.9%. Single nucleotide substitutions were scattered throughout the entire length of the E gene. The majority of these substitutions occurred at the third codon position. No base insertions or deletions were found among the Indian strains. Transitions were more predominant than transversions. The overall transition/transversion bias was 4.54.

Interestingly, the 37 DENV-2 strains of Indian origin segregated into only two different genotypes: the American genotype, comprising DENV-2 from Latin America, the Caribbean and Pacific Islands (including the earliest TT/Trinidad/1953 strain), and the Cosmopolitan genotype, which has a wide geographical distribution, comprising DENV-2 from Australia, China, Africa and South-East Asia (Fig. 1). The mutations observed in the Indian strains of the American and Cosmopolitan genotype, along with other representative strains that show a close relationship to the Indian isolates, are presented in Table 1. Mutations are known to occur in DENV following passage in mice (Bray *et al.*, 1998) and in mammalian cells (Vasilakis *et al.*, 2009). However, in the present study, it was observed that an isolate with a low passage history (NIV 053598) had 12

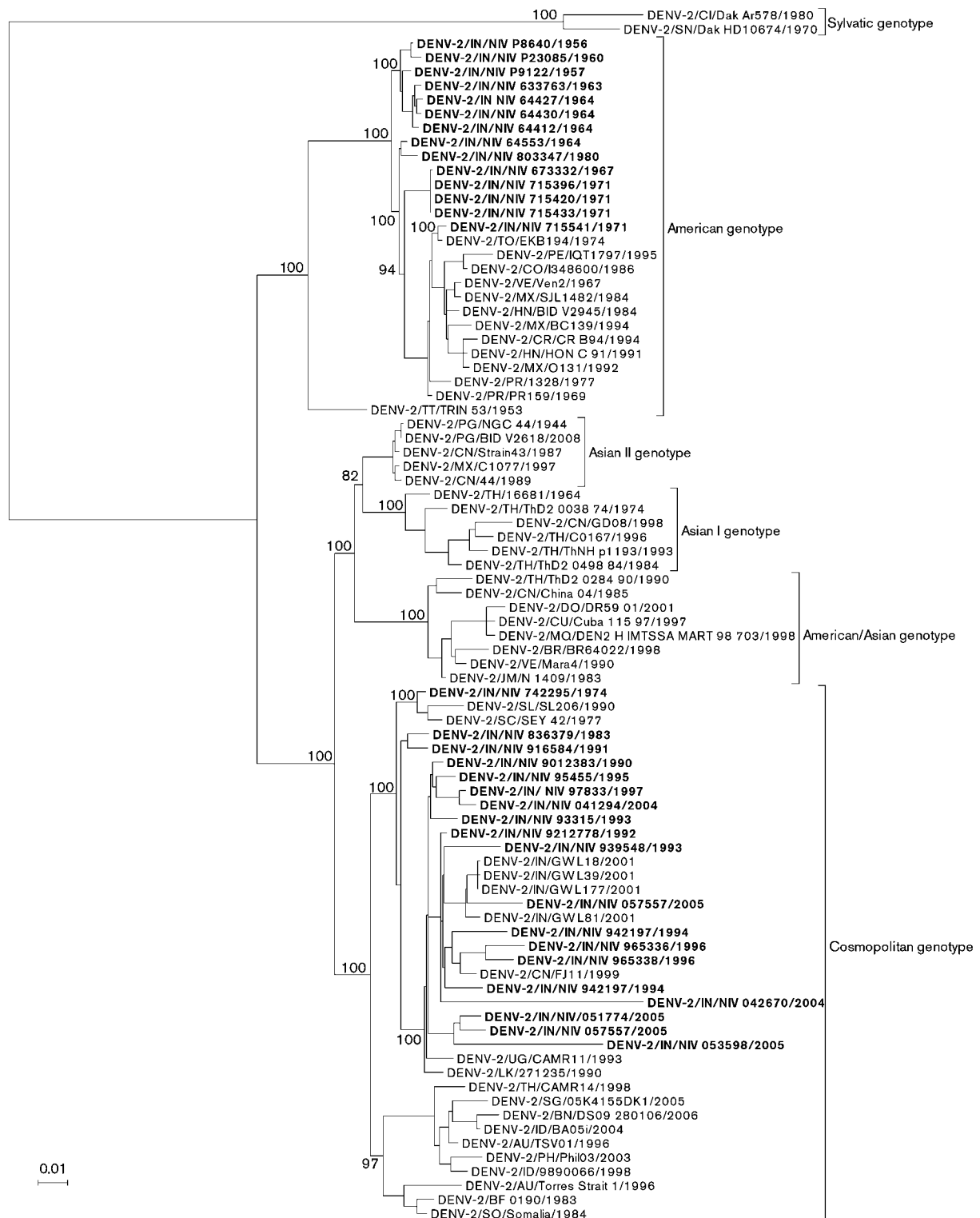


Fig. 1. Phylogenetic tree of DENV-2 based on complete E gene sequences, using maximum likelihood (ML) with PAUP 4b10 software. Indian sequences are shown in bold.

Table 1. Amino acid changes observed for Indian isolates of the American and Cosmopolitan genotypes compared with the prototype New Guinea C strain, along with a few other representative global strains

Unique changes observed in a single strain have been omitted.

Strain	Amino acid position																									
	47	52	57	64	71	81	87	112	124	126	129	134	139	141	149	150	162	164	173	174	191	192	197	200	203	229
DENV-2/PG/NGC44/1944	E	Q	R	K	D	S	D	G	N	K	V	N	I	I	H	A	I	I	A	E	L	D	V	Q	N	S
American genotype																										
DENV-2/TT/TRIN53/1953	K	I	V	D	.
DENV-2/PR/PR159/1969	K	T	.	.	.	E	I	.	V	.	.	.	V	D	.
DENV-2/VE/Ven2/1967	K	T	.	.	.	E	I	.	V	.	.	.	V	D	.
DENV-2/MX/SJL1482/1984	K	T	.	.	.	E	I	.	V	.	.	.	V	D	.
DENV-2/PE/IQT1797/1995	K	T	.	.	.	E	I	.	V	.	.	.	V	D	.
DENV-2/IN/NIV P8640/1956	K	T	.	.	I	.	I	.	V	.	.	.	V	D	.
DENV-2/IN/NIV P9122/1957	K	T	.	.	.	E	I	.	V	.	.	.	V	D	.
DENV-2/IN/NIV 633763/1963	K	T	I	.	V	.	.	.	V	D	.
DENV-2/IN/NIV 64553/1964	K	T	.	.	.	E	I	.	V	.	.	.	V	D	.
DENV-2/IN/NIV 673332/1967	K	T	I	.	V	.	.	.	V	D	.
DENV-2/IN/NIV 715396/1971	K	T	I	.	V	.	.	.	V	D	.
DENV-2/IN/NIV 715541/1971	K	T	.	.	.	E	I	.	V	.	.	.	V	D	.
DENV-2/IN/NIV 803347/1980	K	T	.	A	S	.	I	.	V	.	Y	.	V	D	.
Cosmopolitan genotype																										
DENV-2/SC/SEY 42/1977	K	.	.	.	A	E	N	.	.	V
DENV-2/SL/SL206/1990	K	.	.	.	A	E	N	.	.	V
DENV-2/LK/271235/1990	K	.	.	.	A	E	I	.	.	.	N	.	.	V	L
DENV-2/UG/CAMR11/1993	K	L	.	.	A	E	I	.	.	.	N	.	.	V
DENV-2/CN/FJ11/1999	K	H	.	R	A	E	I	.	.	V	N	.	.	V
DENV-2/IN/NIV 742295/1974	K	.	.	.	A	E	N	.	.	V
DENV-2/IN/NIV 836379/1983	K	.	.	.	A	.	.	S	N	.	.	V
DENV-2/IN/NIV 9012383/1990	K	.	.	.	A	I	.	.	.	N	.	.	V
DENV-2/IN/NIV 9212778/1992	K	.	.	.	A	E	T	.	.	.	N	.	.	V
DENV-2/IN/NIV 916584/1991	K	.	.	.	A	E	I	.	.	.	N	.	.	V
DENV-2/IN/NIV 939548/1993	K	.	.	.	A	I	.	.	V	N	.	.	V	.	.	S	.	A	P	.	.
DENV-2/IN/NIV 93315/1993	K	.	.	.	A	I	.	.	.	N	.	.	V
DENV-2/IN/NIV 942197/1994	K	.	.	.	A	I	.	.	V	N	.	.	V
DENV-2/IN/NIV 942201/1994	K	.	T	.	A	.	H	.	.	.	I	.	.	V	N	.	.	V	.	.	F	A
DENV-2/IN/NIV 95455/1995	K	.	.	.	A	I	.	.	.	N	.	.	V
DENV-2/IN/NIV 965336/1996	K	.	.	.	A	I	.	.	V	N	.	.	V
DENV-2/IN/NIV 965338/1996	K	.	.	.	P	E	I	.	.	V	N	.	.	V
DENV-2/IN/NIV 97833/1997	K	.	.	.	A	I	D	.	.	N	.	.	V
DENV-2/IN/GWL18/2001	K	.	.	.	A	E	I	.	.	V	N	T	.	V
DENV-2/IN/GWL81/2001	K	.	.	.	A	E	I	.	.	V	N	.	.	V
DENV-2/IN/NIV 041294/2004	K	.	.	.	A	E	I	.	.	.	N	.	.	V
DENV-2/IN/NIV 042670/2004	K	.	G	S	E	R	H	.	.	E	I	.	.	.	N	.	.	V	P	Q
DENV-2/IN/NIV 051774/2005	K	.	.	.	A	E	I	I	.	.	N	.	.	V
DENV-2/IN/NIV 053598/2005	K	.	.	.	A	E	I	.	L	.	N	.	.	V	.	A	.	N	.	P	.	.
DENV-2/IN/NIV 057557/2005	K	.	.	.	A	E	I	.	.	.	T	G	.	V	P	.	.	G
DENV-2/IN/NIV 057561/2005	K	.	.	.	A	E	I	.	.	V	N	.	.	V	T

Table 1. cont.

Strain	Amino acid position																			
	272	301	308	309	310	311	314	322	338	346	360	390	393	399	402	404	417	423	444	454
	M	M	V	V	K	E	E	I	E	H	E	N	K	G	I	T	D	G	Y	I
	461	459	462	484	489	491														
	V	L																		
American genotype																				
DENV-2/PG/NGC44/1944	D	.	.	L	.	.	.	T	.
DENV-2/TT/TRIN53/1953	D	.	.	F	.	.	.	T	.
DENV-2/PR/PR159/1969	.	.	I	D	R	.	F	.	.	.	T	.
DENV-2/VE/Ven2/1967	.	.	I	D	R	.	F	.	.	.	T	.
DENV-2/MX/SIL1482/1984	.	.	I	D	.	.	F	.	.	.	T	.
DENV-2/PE/IQT1797/1995	D	.	.	F	.	.	.	T	.
DENV-2/IN/NIV P8640/1956	D	.	.	L	.	.	.	T	T
DENV-2/IN/NIV P9122/1957	D	.	.	F	.	.	R	T	.
DENV-2/IN/NIV 633763/1963	D	.	.	F	.	.	.	T	.
DENV-2/IN/NIV 64553/1964	D	.	.	F	.	.	.	T	.
DENV-2/IN/NIV 673332/1967	K	D	.	.	F	N	.	.	T	.
DENV-2/IN/NIV 715396/1971	K	D	.	.	F	N	.	.	T	.
DENV-2/IN/NIV 715541/1971	.	.	I	D	.	.	F	.	.	.	T	.
DENV-2/IN/NIV 803347/1980	D	.	.	L	.	.	.	T	.
Cosmopolitan genotype																				
DENV-2/SC/SEY 42/1977	S	.	.	F	.	.	.	T	.
DENV-2/SL/SL206/1990	S	.	.	F	.	.	.	T	.
DENV-2/LK/271235/1990	S	.	.	F	.	.	.	T	.
DENV-2/UG/CAMR11/1993	Y	.	S	.	.	F	.	.	G	T	.
DENV-2/CN/F11/1999	S	.	.	F	.	.	.	T	.
DENV-2/IN/NIV 742295/1974	S	.	.	V	.	.	.	T	.
DENV-2/IN/NIV 836379/1983	S	.	.	F	.	.	.	T	.
DENV-2/IN/NIV 9012383/1990	S	.	.	F	.	.	.	T	.
DENV-2/IN/NIV 9212778/1992	F	.	.	.	T	.
DENV-2/IN/NIV 916584/1991	V	F	.	.	.	T	.
DENV-2/IN/NIV 939548/1993	R	S	.	.	F	.	.	.	T	.
DENV-2/IN/NIV 93315/1993	S	.	S	F	.	.	.	T	.
DENV-2/IN/NIV 942197/1994	.	I	.	G	R	.	A	S	.	.	F	.	C	R	.	.
DENV-2/IN/NIV 942201/1994	L	.	S	.	R	F	.	.	.	T	.
DENV-2/IN/NIV 95455/1995	S	.	.	F	.	.	.	T	.
DENV-2/IN/NIV 965336/1996	V	L	.	G	.	.	.	V	.	.	.	S	.	.	F	.	.	.	S	C
DENV-2/IN/NIV 965338/1996	V	.	.	G	S	.	.	F	.	.	.	T	M
DENV-2/IN/NIV 97833/1997	S	.	.	F	.	.	.	T	.
DENV-2/IN/GWL18/2001	V	.	.	.	S	.	.	F	.	.	.	T	.
DENV-2/IN/GWL81/2001	V	.	.	.	S	.	.	F	.	.	.	T	.
DENV-2/IN/NIV 041294/2004	S	.	.	F	.	.	.	T	.
DENV-2/IN/NIV 042670/2004	W	.	.	S	.	.	F	P	.	.	T	.
DENV-2/IN/NIV 051774/2005	S	.	.	F	.	.	.	T	.
DENV-2/IN/NIV 053598/2005	W	.	.	H	N	.	G	T	Q	.	F	.	.	.	T	.
DENV-2/IN/NIV 057557/2005	G	S	.	.	F	.	.	.	T	.
DENV-2/IN/NIV 057561/2005	K	V	.	.	G	S	.	.	F	.	.	.	T	.

unique mutations, whereas an isolate with a high passage history (NIV 95455) revealed only one unique mutation.

DENV-2 strains that were isolated before and during 1971 (pre-1971 Indian isolates) grouped within the American genotype, whilst those isolated after 1974 (post-1971 Indian isolates) grouped within the Cosmopolitan genotype cluster. An exception was strain IN/NIV803347/1980, isolated during the DENV-2 epidemic in Kolkata (West Bengal) in 1980, which grouped within the American genotype cluster. This strain was seen to possess several of the mutations specific to the American genotype, i.e. 81T, 139V, 162V, 203D, 390D and 484I (Table 1). The PNI of the pre-1971 Indian isolates with respect to strain PG/NGC44/1944 was 91.2–92.3 %, whilst that of the post-1971 Indian isolates was 89.0–95.4 %.

Within the American genotype, the Indian isolates formed four separate clusters with 100 % bootstrap support. In the fourth group, the IN/NIV715541/1971 isolate was found to be related most closely to strain TO/EKB194/1974 (isolated from Tonga) and these were related most closely to the native American isolates from South America from 1967 to 1995. Several of these isolates share a mutation (V308I) specifically with the IN/NIV715541/1971 strain. Within the Cosmopolitan group of viruses, two clusters with strong bootstrap support were observed. The Indian strains were distinct from the other Cosmopolitan strains. Further, within the Indian cluster, two main groups were observed. The Indian strain from 1974 (IN/NIV742295/1974) was related closely to the SC/SEY42/1977 and SL/SL206/1990 strains from the Seychelles and Sierra Leone, respectively. The group of Indian strains from 1994, 1996, 2001 and 2005, one strain from 1993 and one from 2004 showed a close relationship with strains from Sri Lanka (LK/271235/1990), Uganda (UG/CAMR11/1993) and China (CN/FJ11/1999). These isolates shared a mutation Lys126→Glu, whilst strain CN/FJ11/1999 specifically shared a mutation of Ile141→Val with most of the Indian Cosmopolitan strains. Of the six amino acid replacements (Glu71→Ala, Ile129→Val, His149→Asn, Ile164→Val, Asn390→Ser and Ile462→Val) noted in the viruses of the Cosmopolitan genotype (Twiddy *et al.*, 2002a), five were seen in the Indian viruses at positions 71, 149, 164, 390 and 462. At position 129, most Indian viruses retained Ile, except for two strains belonging to the Cosmopolitan genotype that showed either Val (as in the ancestral prototype strain PG/NGC44/1944) or Thr. At position 390, two Indian Cosmopolitan strains possessed Asn, as in strain PG/NGC44/1944. In addition, five of eight Indian strains belonging to the American genotype possessed Lys at position 126, as in the prototype strain PG/NGC44/1944, whereas others possessed Glu. Notably, at the same position, 12 of 21 Indian strains of the Cosmopolitan genotype possessed Glu, whereas the rest had Lys. At position 322, six of 21 Indian strains of the Cosmopolitan genotype possessed Val instead of Ile, which was seen in the rest of the Indian

strains. Furthermore, at position 402, most Indian strains possessed Phe, although a minority (two of eight) of Indian strains belonging to the American genotype possessed Leu and one strain from the Cosmopolitan genotype possessed Val at that position.

Selection pressure analysis

Site-specific selection pressure was analysed by using likelihood procedures including single-likelihood ancestor counting (SLAC), random-effects likelihood (REL) and fixed effects likelihood (FEL). Sites were considered to be under positive selection if at least two of the methods indicated this with high statistical significance ($P < 0.1$ /Bayes factor > 50). Different datasets were used, one consisting of all American genotype sequences ($n = 21$) and one consisting of all Cosmopolitan genotype sequences ($n = 36$). Within the American genotype strains, site 402, and within the Cosmopolitan genotype, site 126, satisfied the criteria (with high significance). In the Cosmopolitan genotype, three other sites, 64, 310 and 360, showed some evidence of positive selection (Table 2). We also considered two other datasets by splitting the Indian sequences into the American genotype ($n = 8$) and the Cosmopolitan genotype ($n = 21$) to examine whether similar selection pressure processes could be identified. We found that site 402 in the American genotype and site 126 in the Cosmopolitan genotype were the only ones to be under significantly high positive selection by at least two of the likelihood methods adopted.

Molecular clock analysis

Estimates of the nucleotide substitution rates and the time to most recent common ancestor (tMRCA) of the DENV-2 genotypes as well as Indian DENV-2 strains were inferred from a total of 76 (29 Indian and 47 global) dated E gene sequences. Among all the models (strict, relaxed uncorrelated exponential and relaxed uncorrelated lognormal) of molecular clocks employed, models enforcing relaxed molecular clocks performed better than the strict clock model (Table 3). The logistic growth demographic model did not converge under strict and relaxed (uncorrelated exponential and uncorrelated lognormal) models of molecular clock. Also, the expansion growth demographic model under relaxed (uncorrelated exponential and uncorrelated lognormal) models and the exponential growth demographic model under a relaxed (uncorrelated lognormal) model of molecular clock did not converge with the dataset considered.

The highest marginal likelihood was obtained with the model implementing the relaxed uncorrelated exponential molecular clock and the constant population growth model. The Bayes factor and the Bayesian skyline plot (data not shown) gave stronger support to this model than other clock and population models. Under this model, the mean substitution rate was 6.5×10^{-4} substitutions per site

Table 2. Parameters from selection pressure analysis of the E gene using the SLAC, REL and FEL methods, based on different datasets as indicated

Sites under strong positive selection pressure as demonstrated by at least two methods are shown in bold. Other sites that are indicated to be under positive selection by only one of the methods, with the other two methods indicating close to threshold *P* values, are italicized.

Codon	SLAC		FEL		REL	
	<i>dN/dS</i>	<i>P</i> value*	<i>dN/dS</i>	<i>P</i> value*	<i>dN/dS</i>	Bayes factor†
American genotype (<i>n</i> =21)						
124	−0.71	0.77	5.42	0.71	0.58	78.08
126	1.60	0.67	5.35	0.86	0.70	4209.02
402	12.45	0.22	15.64	0.05	1.00	62745.60
404	4.78	0.51	−0.42	0.98	0.69	76.17
Cosmopolitan genotype (<i>n</i> =36)						
64	2.51	0.39	1.82	0.13	0.45	123.58
126	5.74	0.13	4.63	0.05	0.56	642.31
129	−1.41	0.78	−0.31	0.86	0.13	98.26
310	2.54	0.40	1.92	0.12	0.47	139.12
314	0.88	0.62	0.34	0.89	0.30	171.96
322	3.09	0.34	1.27	0.49	0.33	270.58
338	−1.48	0.81	−0.63	0.80	0.11	86.11
360	4.13	0.22	3.20	0.10	0.56	650.07
390	1.82	0.49	1.64	0.34	0.45	494.35
417	0.99	0.59	1.15	0.50	0.42	335.23
461	−0.89	0.79	−0.01	0.99	0.06	50.85
Indian strains (American genotype, <i>n</i> =8)						
124	27.93	0.51	153.61	0.14	4.40	1079.60
126	27.51	0.55	192.93	0.26	4.37	847.87
402	46.39	0.27	146.81	0.09	4.45	1674.43
404	22.29	0.52	51.74	0.85	4.30	589.41
Indian strains (Cosmopolitan genotype, <i>n</i> =21)						
126	8.95	0.19	11.42	0.05	2.68	991.00
461	4.91	0.29	3.11	0.07	0.15	3.82

*A *P* value of <0.1 indicates statistically significant evidence for positive selection.

†A Bayes factor of >50 indicates statistically significant evidence for positive selection.

year^{−1} [95 % highest probability density (HPD), 4.1–8.7 × 10^{−4}] (Table 3). The maximum clade credibility tree generated under the best fit uncorrelated exponential molecular clock with constant population growth model is shown in Fig. 2.

The estimate of the tMRCA for all genotypes (excluding the Sylvatic genotype) of DENV-2 was about 200 years (95 % HPD: 90, 377 years) (Fig. 2). Strain TT/Trinidad/1953 and the American genotype strains share a common ancestor about 106 years (95 % HPD: 61, 185 years) ago, whilst all of the pre-1971 Indian isolates with the American genotype shared a common ancestor (node A) about 73 years (95 % HPD: 54, 100 years) ago. Three subgroups of Indian isolates with high bootstrap support were observed. The first group of Indian isolates from 1956 to 1963 showed a tMRCA of about 60 years (95 % HPD: 52, 72 years), whilst the second group of two Indian isolates from 1964 and 1980 showed a tMRCA of

about 51 years (95 % HPD: 44, 62 years) and diverged from node B about 62 years (95 % HPD: 50, 79 years) ago. The third group of Indian isolates (1967–1971), along with the native American isolates, shared a common ancestor (node C) about 57 years (95 % HPD: 47, 69 years) ago.

The common ancestor of all of the Cosmopolitan isolates was about 65 years (95 % HPD: 41, 101 years) old. The subclade consisting of east African and South-East Asian strains showed a tMRCA of about 43 years (95 % HPD: 30, 60 years), whilst that of the post-1971 Indian isolates was about 47 years (95 % HPD: 37, 63 years). Within the Indian subclade, the two subgroups indicated similar tMRCAs of about 39 years (95 % HPDs: 29, 50, and 34, 46 years, respectively) ago (nodes D and E).

The tMRCA of the Asian I genotype was about 58 years (95 % HPD: 45, 79 years), that of the Asian II genotype was

Table 3. Parameter estimates for different clock models for E gene sequences of DENV-2 using BEAST

The best fit model is indicated in bold.

Clock model	Posterior	Marginal likelihood	Mean rate $\times 10^{-4}$ (95 % HPD) (subs per site year ⁻¹)	Root age (95 % HPD) (years)	tMRCA (95 % HPD) (years)	
					American genotype	Cosmopolitan genotype
Strict clock, constant size	-10570 (-10550, -10590)	-10084.7 (± 0.13)	5.0 (4.0, 6.0)	156.7 (130.1, 184.1)	111.8 (94.3, 130.2)	81.6 (65.9, 97.7)
Strict clock, exponential growth	-10560 (-10540, -10580)	-10081.1 (± 0.13)	4.8 (3.8, 5.8)	152.6 (128.3, 179.6)	110.0 (92.9, 127.8)	82.8 (67.6, 99.2)
Strict clock, expansion growth	-10560 (-10540, -10580)	-10080.4 (± 0.14)	4.7 (3.7, 5.7)	160.5 (133.3, 190.5)	113.4 (95.5, 132.6)	83.4 (68.1, 100.8)
Uncorrelated exponential, constant size	-10310 (-10280, -10340)	-9887.1 (± 0.18)	6.5 (4.1, 8.7)	200.4 (89.8, 376.6)	105.86 (60.9, 185.4)	64.49 (40.7, 101.0)
Uncorrelated exponential, exponential growth	-10310 (-10280, -10340)	-9888.1 (± 0.17)	7.1 (5.2, 9.0)	128.0 (83.0, 186.7)	84.5 (60.9, 115.8)	62.0 (41.5, 91.0)
Uncorrelated lognormal, constant size	-10340 (-10310, -10370)	-9899.1 (± 0.18)	6.09 (4.4, 7.7)	178.6 (104.9, 273.9)	118.7 (70.8, 180.0)	71.9 (46.7, 101.1)

about 75 years (95 % HPD: 65, 93 years) and that of the American/Asian genotype was fairly recent at about 39 years (95 % HPD: 27, 58 years).

DISCUSSION

The number of different genotypes seen in each serotype is consistent with the abundance of genetic diversity in DENV (Holmes & Burch, 2000). It is important to note that, among the four serotypes, serotype 2 has been the most genetically diverse (Rico-Hesse *et al.*, 1997). Furthermore, it has been observed that some of the genotypes have a wide geographical distribution (Twiddy *et al.*, 2002a, b; Zhang *et al.*, 2005). Various earlier studies have shown six distinct genotypes of DENV-2 (Rico-Hesse, 1990; Wang *et al.*, 2000; Twiddy *et al.*, 2002a).

Our phylogenetic analysis based on the E gene of DENV-2 strains isolated during the period from 1956 to 2005 in India, along with globally representative strains, showed that Indian strains segregated into two genotypes, namely the American and Cosmopolitan genotypes. The Indian isolates obtained before and during the 1971 epidemics clustered into the American genotype (Fig. 1), whereas the post-1971 isolates clustered into the Cosmopolitan genotype. There was no indication of any geographical dispersal pattern. Our molecular clock analysis (Fig. 2) showed that the tMRCA of the pre-1971 Indian strains and the American genotype was about 73 years, with an upper bound of 100 years. This time period of 1905–1932 corresponds with the Indian immigration to Surinam and Guyana between 1873 and 1916 (Chickrie, 2003) and suggests transportation of the virus from the Americas to India. Unfortunately, the lack of samples from earlier collection times does not allow for more robust analysis.

Figs 1 and 2 also show the different subgroups among the pre-1971 Indian isolates; these indicate the repeated establishment of the genotype at least three times in India (nodes A, B and C), the mean divergence times based on the molecular clock analysis being 73 years (about 1932), 62 years (1943) and 57 years (1948), respectively. The tree in Fig. 2 shows clearly that the American strains during the period from 1967 to 1995 share a common ancestor with the Indian strains of the period from 1967 to 1971 at node C. This is indicative of a possible transmission of DENV-2 viruses between India and the Americas sometime around 1948. Incidentally, these dates coincide with World War II and its aftermath, during which global movements showed an increase.

The Indian strains isolated during and after 1974 segregated into the Cosmopolitan genotype, suggesting clearly that there was a change in the circulating genotype of DENV-2 in India. A deviation from this pattern was observed with the 1980 strain (IN/NIV803347/1980) isolated from Kolkata (West Bengal), which belonged to the American genotype. The close similarity of this isolate to the Indian isolate of 1964 (NIV_64553) may, however,

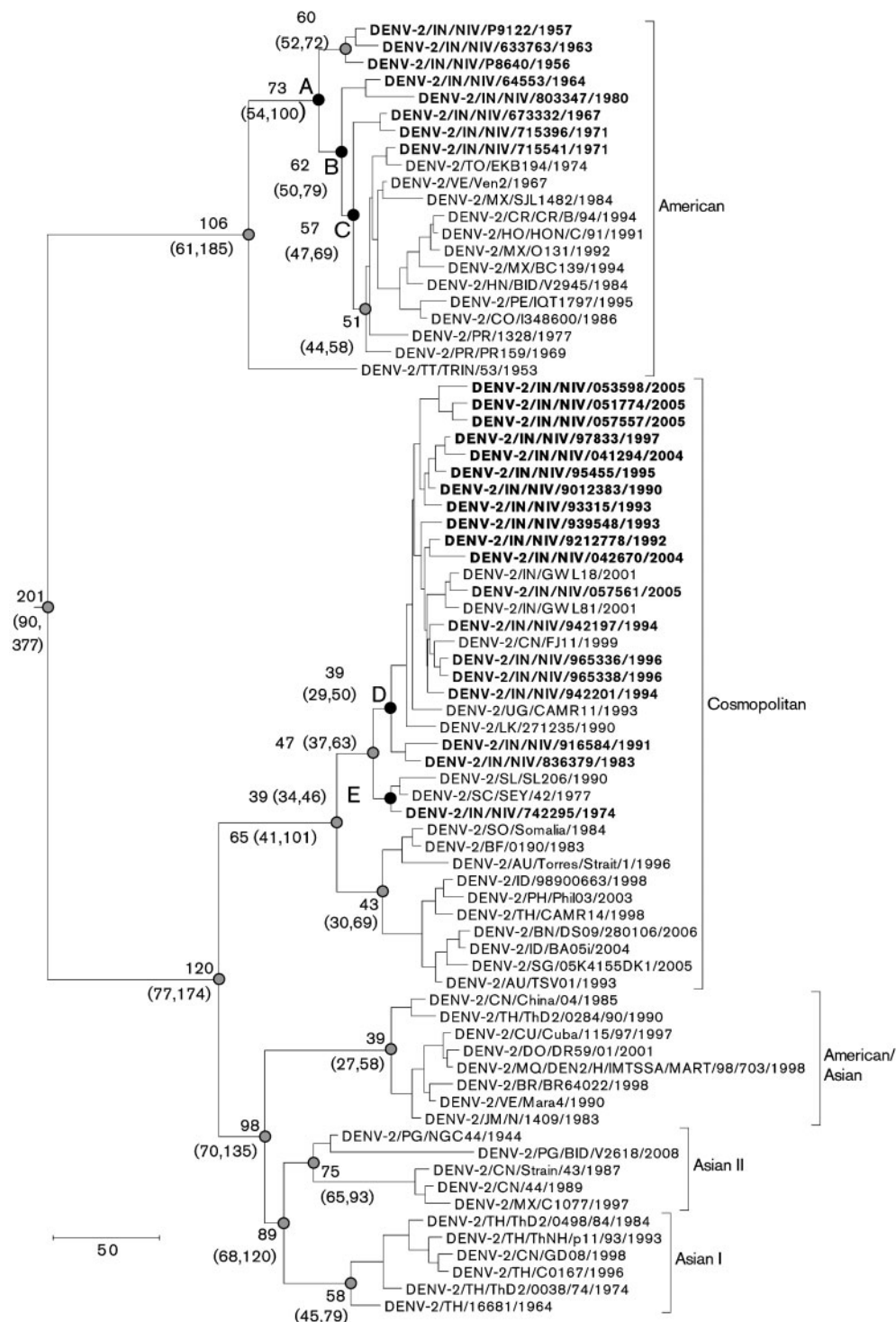


Fig. 2. Bayesian Markov chain Monte Carlo (MCMC) tree of E gene sequences of DENV-2 with the constant size demographic model under the relaxed uncorrelated exponential clock model. The tMCA estimates of the key nodes (highlighted by black and grey circles) are also shown; black circles (labelled A–E) indicate nodes representing divergence of Indian subgroups. Indian sequences are shown in bold.

imply that, rather than it being an imported case, the American genotype was circulating during that time. The Cosmopolitan genotype seems to have been established in India in the post-1980 period.

An earlier study of genomic sequence analysis of DENV-2 from the 1996 epidemic in Delhi, India (Singh *et al.*, 1999), showed that the replacing Cosmopolitan genotype resembled strains of the Somalia and Torres Straits. Our phylogenetic analysis of E gene sequences showed that the Cosmopolitan genotype segregated into two subclades with 100% bootstrap support. The subclade consisting of the post-1971 Indian isolates was distinct from the subclade consisting of the Somalia and Torres Strait strains [along with later strains from South-East Asia (specifically Indonesia, the Philippines, Taiwan, Singapore and Malaysia)], indicating that there have may been independent introductions in the two regions. The post-1971 Indian isolates formed two subgroups (nodes D and E in Fig. 2) with strong bootstrap support (100 and 99%, respectively), indicative of two establishments of the Cosmopolitan genotypes in India. In one of these subgroups, the earliest Indian isolate (IN/NIV742295/1974) shared a common ancestor with SC/SEY42/1977 and SL/SL206/1990, implying dispersal of a strain from Trichur, Kerala (IN/NIV742295/1974) to the island in the Indian Ocean and further spread to the West African countries. The Indian isolates of the period from 1983 to 1991 (IN/NIV836379/1983 and IN/NIV916584/1991) grouped with strains from Sri Lanka (LK/271235/1990) and Uganda (UG/CAMR11/1993), indicating dispersal of the Indian strains to these countries. Notably, a strain from Fujian province, China (CN/FJ11/1999), also fell into the cluster of Indian isolates. Close relatedness of this strain to the Indian isolates (IN/NIV965336/1996 and IN/NIV965338/1996) of an earlier timescale is indicative of the possible introduction of this particular strain to China from India. The transfer of viruses across continents and oceans can be attributed to increased global movements, as well as transportation of infected mosquitoes via commercial shipping (Self, 1984; Song *et al.*, 2003; Zeller, 1998).

The major mutations unique to the subclade defining the Indian strains were identified in the sequence-based analyses carried out in this study. The tMRCA of the group consisting of all post-1971 Indian isolates was about 47 years, with an upper bound of about 63 years, while that of the two subgroups (nodes D and E) was about 39 years and not more than 50 years, implying migration of DENV-2 to countries across the Indian Ocean during 1955–1966.

The E protein is the major antigenic determinant of DENV. The selection pressure analysis based on the E gene provided strong evidence for positive selection at two previously unreported sites, one (402) in the American genotype and one (126) in the Cosmopolitan genotype. The other sites (64, 310 and 360 in the Cosmopolitan genotype) that showed some evidence of positive selection

have also not been reported previously. Sites Ile402 in the American genotype and Lys126 in the Cosmopolitan genotype were the only sites found to be under strong positive selection among the Indian group of pre- and post-1971 isolates. The American genotype has previously been reported to be under conservative evolution, whilst there has been evidence of diversifying selection in the Cosmopolitan genotype (Twiddy *et al.*, 2002a). Among the sites reported in the Cosmopolitan genotype (positions 52, 159 and 390) in the E gene, none were identified to be under positive selection with high statistical significance in our study. Most of these sites were found to be invariable in the Indian context and hence may have ceased to show evidence of any selection pressure.

Notably, site 402 showed evidence of strong positive selection in the American genotype among both global and Indian viruses. This site has been proposed to be located within the postulated stem-anchor region adjacent to the putative receptor-binding domain (domain III) of the E protein structure. Substitutions at this position were reported to confer neurovirulence in mice (Bray *et al.*, 1998). Site 126, identified in this study to be under selection pressure in the Cosmopolitan genotype, was found to be under mutational pressure during the clade replacement. The amino acid residue at position 126 resides in domain II of the E glycoprotein (Rey *et al.*, 1995). It has been shown to be responsible for a conformational change at low pH and to be involved in exposing the fusion peptide on the virion surface at the time of virus entry into susceptible cells. This residue could, therefore, be responsible for altering virus–cell interactions (Roehrig *et al.*, 1994). Substitutions altering virus–cell interactions may be one of several factors responsible for broad dispersal of the Cosmopolitan genotype. In the case of the Indian viruses, these substitutions may provide a fitness advantage over the American genotype, which led to detected change in DENV-2 epidemiology during the past few decades. The functional significance of other sites, such as 64, 310 and 360, that were also indicated to be under selection pressure, in terms of viral fitness, is not clear, but may merit further study.

In earlier studies, the role of natural selection pressure has also been described for strain extinction and/or replacement (Sittisombut *et al.*, 1997; Wittke *et al.*, 2002). Further, strain replacement and/or extinction has also been attributed to stochastic processes, such as changes in the population density and the vector mosquito density during inter-epidemic years, or is a regular occurrence in virus evolution, especially at times when the numbers of susceptible hosts or mosquitoes are low (Holmes & Twiddy, 2003; Vasilakis *et al.*, 2007). In the Indian scenario, it is not possible to elucidate whether the genotype replacement is attributed to the fitness of the latter genotype or because of stochastic processes. Based on the findings reported here, however, it seems reasonable to suggest that positive selection may have some role to play in genotype replacement.

The estimated rate of nucleotide substitution in the E gene of the Indian strains of DENV-2 was shown to be about 6.5×10^{-4} substitutions per site year⁻¹. This average rate is comparable to the earlier estimates of roughly 6×10^{-4} substitutions per site year⁻¹, considering global DENV-2 human strains (Twiddy *et al.*, 2003; Wang *et al.*, 2000). The overall time of emergence of human DENV-2 was estimated to be about 200 years (95 % HPD: 90, 377 years) ago, which is comparable to the earlier estimates of Zanotto *et al.* (1996) and Wang *et al.* (2000). The estimate of the divergence time of the American, Asian, Cosmopolitan and American/Asian genotypes was found to be about 106 (61, 185), 90 (68, 120), 65 (41, 101) and 39 (27, 58) years, respectively. These estimates are also comparable with the earlier estimates (Twiddy *et al.*, 2003; Wang *et al.*, 2000).

Notably, the Americas experienced a rapid displacement of the less virulent American genotype by the more virulent Asian and American/Asian genotypes (Rico-Hesse *et al.*, 1997; Leitmeyer *et al.*, 1999). Whilst the Asian strains from Thailand, Indonesia and China may have moved westward and displaced the American genotype, it is noteworthy that these genotypes were never established in India, at least on the basis of the available genetic data.

Overall, the present study shows a change in the circulating genotype of DENV-2 in India from the American to the Cosmopolitan genotype during the 1970s. The selection pressure analysis on E gene sequences suggests the presence of positive selection at two sites, Ile402 in the American genotype (as well as in Indian pre-1971 strains) and Lys126 in the Cosmopolitan genotype (as well as in Indian post-1971 strains). The rates of nucleotide substitution determined in this study are comparable with the earlier reports suggesting similar rates of substitution for the Indian strains. The tMRCA of the pre-1971 Indian strains with the American genotype was about 73 years and not more than 100 years ago (1905–1932), correlating with historical records of traffic between India and South America. The tMRCA of the Cosmopolitan genotype, including all post-1971 Indian isolates, was about 47 years, with estimates of transmission of DENV-2 to countries across the Indian Ocean about 39 years and not more than 50 years ago, i.e. between 1955 and 1966. Although the present study helps us to understand the establishment and dispersal of the two genotypes in the Indian context, analysis based on full genome sequences of the isolates would consolidate the present conclusions.

METHODS

Viruses. The DENV-2 strains used in this study were isolated from different parts of India during different epidemics spanning over almost 50 years (1956–2005). The earliest isolates, from between 1956 and 1964, were from Vellore district, Tamil Nadu, south India, while the rest of the isolates were from different regions covering north, west, central and south India. All strains except for two (IN/NIV715541/1971 and IN/NIV051774/2005) were isolated from

human samples; these two strains were isolated from *A. aegypti* mosquitoes in 1971 and 2005, respectively. Thirty-three isolates were sequenced as part of this study (details of these 33 virus strains are presented in Supplementary Table S1, available in JGV Online); the other four sequences were retrieved from GenBank. Viruses were procured from the virus repository of the National Institute of Virology, Pune, India. Lyophilized viruses were reconstituted in sterile distilled water and inoculated into Swiss albino suckling mice by the intracerebral route. Mice were maintained as per the guidelines of the Committee for Protection, Supervision and Control of Experiments on Animals (CPSCEA). Mice were observed for sickness; sick mice were euthanized and their brains were harvested. A 10 % suspension of mouse brain in 0.75 % bovine albumin phosphate saline (BAPS) was prepared and stored at -70°C until use.

RNA extraction, RT-PCR and sequencing. Infectious mouse brain suspension representing each strain was used for RNA extraction. Viral RNA was isolated by using a QIAamp Viral RNA Mini kit (Qiagen), according to the manufacturer's instructions. cDNAs were prepared with 10 μl RNA template using avian Moloney virus reverse transcriptase (Promega) as per the manufacturer's instructions. The E gene was amplified by using primers described by Wang *et al.* (2000) (see Supplementary Table S2, available in JGV Online) with 2.0 μl cDNA by using Platinum *Taq* polymerase (Invitrogen). PCR products were visualized by ethidium bromide agarose gel staining and the PCR products were gel-purified by using a QIAquick gel extraction kit (Qiagen). The purified PCR products were sequenced by using a BigDye Terminator cycle sequencing ready reaction kit (Applied Biosystems) on an automatic sequencer (ABI PRISM Genetic Analyzer 3100; Applied Biosystems).

Sequence and phylogenetic analysis. Multiple sequence alignment was done by using CLUSTAL W implemented in MEGA v. 3.1 (Kumar *et al.*, 2004). A phylogenetic tree was constructed by using the E gene sequences of the 33 strains (Supplementary Table S1), along with 51 other DENV-2 sequences from different geographical regions obtained from GenBank (see Supplementary Table S3, available in JGV Online). In the process of global sequence selection, sequences that were related closely (>99.5 % identity) were eliminated. Two sequences of DENV-2 Sylvatic strains were used as outgroup sequences. The tree, based on 86 sequences, was constructed by the maximum-likelihood (ML) method in the software PAUP 4b10 (Swofford, 2003) by using a heuristic search algorithm in two steps. The substitution model used was the best model as detected by the Akaike Information Criterion (AIC) in MODELTEST 3.7 (Posada & Crandall, 1998), namely GTR + G4 + I (general time-reversible model with gamma-distributed rates of variation among sites and a proportion of invariable sites). In the first step, branch-swapping by nearest neighbour interchange was used with stepwise, random addition of sequences. In the second step, the tree bisection–reconnection method of branch-swapping was applied on the best tree found in the first step. The ML estimates of different parameters of the nucleotide substitution model were as follows: the optimal base frequencies of A, C, G and T were 0.33, 0.21, 0.24 and 0.20, respectively. The estimates of the relative substitution rates were 0.89, 6.41, 1.41, 1.1, 18.71 and 1.0 for the substitution types A \leftrightarrow C, A \leftrightarrow G, A \leftrightarrow T, C \leftrightarrow G, C \leftrightarrow T and G \leftrightarrow T, respectively. The estimated proportion of invariable sites was 24.6 %, whilst the shape parameter of the Γ (gamma) distribution of rate variation among sites was 0.87.

The topology as predicted by the ML tree was also confirmed by two more methods: maximum parsimony (MP) and the Bayesian MCMC (Markov chain Monte Carlo) tree-sampling method. The MP tree construction was done by a heuristic search approach implemented in PAUP 4b10. MrBayes v. 3 (Ronquist & Huelsenbeck, 2003) was used to build the Bayesian MCMC tree using GTR + G4 + I as the nucleotide substitution model with 1 000 000 generations, a sampling frequency

of 100 and 25 % of the generations as burn-in. The reliability of the nodes of the ML tree was judged by superimposing the posterior supports (%) at corresponding nodes of the Bayesian MCMC tree.

Selection pressure analysis. To identify the existence of positive selection pressure at individual codon sites in the E gene of DENV-2, three likelihood procedures were used: the SLAC and FEL methods and the more powerful REL method (Pond & Frost, 2005). The strength of selection pressure is determined on the basis of the ratio of non-synonymous (dN) to synonymous (dS) substitutions per site (ratio dN/dS). The analysis was carried out using the online Datamonkey facility (<http://www.datamonkey.org/>), incorporating the GTR model of nucleotide substitution, with a phylogenetic tree inferred by using the neighbour-joining method. Four datasets were used: all American genotype, all Cosmopolitan genotype, only Indian sequences of American genotype and only Indian sequences of Cosmopolitan genotype. In all of these datasets, closely related sequences (>99.5 % nucleotide identity) were eliminated. With this restriction, $n=21$, 36, 8 and 21, respectively.

Molecular clock analysis. The dataset for the molecular clock analysis comprised Indian and global sequences from the five different genotypes (Sylvatic strains excluded). No evidence of recombination was identified in the Indian strains or between the Indian and the globally representative viruses. Closely similar sequences (>99.5 % identity) were removed from the analysis. With this restriction, the number of Indian isolates was reduced to 29 and the total dataset consisted of 76 sequences. The rate of nucleotide substitution and divergence times, i.e. tMRCA, of DENV-2 strains of Indian origin and also of the different genotypes were calculated by using the Bayesian MCMC approach as implemented in BEAST 1.4.8 (Drummond & Rambaut, 2007). We employed both strict and relaxed (uncorrelated exponential and uncorrelated lognormal) molecular clocks (Drummond *et al.*, 2006) with different demographic models (constant size, exponential growth, logistic growth and expansion growth). MODELTEST was used to select the best-fitting nucleotide substitution model, using AIC as implemented in MODELTEST 3.7 (Posada & Crandall, 1998). The GTR + G4 + I model was found to be the best fit for our dataset. Five independent MCMC analyses, each for 10 000 000 steps, were performed for each combination of branch rate and demographic model, and combined with a burn-in value set to 10 % generations using LogCombiner 1.4.8 (implemented in BEAST). The resulting convergence was analysed by using Tracer 1.4 (Rambaut & Drummond, 2007). Model comparison was done by calculation of Bayes factor based on the relative marginal likelihoods of the models under comparison (Suchard *et al.*, 2001). The maximum clade credibility tree was generated by using Tree Annotator (available in BEAST), and FigTree 1.2.2 (<http://tree.bio.ed.ac.uk/>) was used for visualization of the annotated trees. The 95 % HPD intervals were obtained to ascertain the uncertainty in the parameter estimates.

ACKNOWLEDGEMENTS

The authors appreciate the input from two anonymous reviewers, which has significantly improved the manuscript. S. R. P. K. acknowledges the Council of Scientific and Industrial Research (Government of India) for providing fellowship to carry out this research. The authors thank Mrs S. S. Athawale and Dr A. B. Sudeep for their help with virus propagation.

REFERENCES

- Arunachalam, N., Murty, U. S., Kabilan, L., Balasubramanian, A., Thenmozhi, V., Narahari, D., Ravi, A. & Satyanarayana, K. (2004). Studies on dengue in rural areas of Kurnool District, Andhra Pradesh, India. *J Am Mosq Control Assoc* **20**, 87–90.
- Bennett, S. N., Holmes, E. C., Chirivella, M., Rodriguez, D. M., Beltran, M., Vorndam, V., Gubler, D. J. & McMillan, W. O. (2006). Molecular evolution of dengue 2 virus in Puerto Rico: positive selection in the viral envelope accompanies clade reintroduction. *J Gen Virol* **87**, 885–893.
- Bray, M., Ruhe, M., Tokimatsu, I. & Lai, C. J. (1998). Genetic determinants responsible for acquisition of dengue type 2 virus mouse neurovirulence. *J Virol* **72**, 1647–1651.
- Chambers, T. J., Hahn, C. S., Galler, R. & Rice, C. M. (1990). Flavivirus genome organization, expression and replication. *Annu Rev Microbiol* **44**, 649–688.
- Chareonsook, O., Foy, H. M., Teeraratkul, A. & Silarug, N. (1999). Changing epidemiology of dengue hemorrhagic fever in Thailand. *Epidemiol Infect* **122**, 161–166.
- Chaturvedi, U. C. (2006). The curse of dengue. *Indian J Med Res* **124**, 467–470.
- Chaturvedi, U. C. & Shrivastava, R. (2004). Dengue hemorrhagic fever: a global challenge. *Indian J Med Microbiol* **22**, 5–6.
- Chickrie, R. (2003). The Afghan Muslims of Guyana and Suriname. http://www.guyana.org/features/afghanguyanese_muslim.html
- Cologna, R. & Rico-Hesse, R. (2003). American genotype structures decrease dengue virus output from human monocytes and dendritic cells. *J Virol* **77**, 3929–3938.
- Cologna, R., Armstrong, P. M. & Rico-Hesse, R. (2005). Selection for virulent dengue viruses occurs in humans and mosquitoes. *J Virol* **79**, 853–859.
- Dar, L., Broor, S., Sengupta, S., Xess, I. & Seth, P. (1999). The first major outbreak of dengue haemorrhagic fever in Delhi, India. *Emerg Infect Dis* **5**, 589–590.
- 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**, 314–322.
- Diallo, M., Ba, Y., Faye, O., Soumare, M., Dia, I. & Sall, A. (2008). Vector competence of *Aedes aegypti* populations from Senegal for sylvatic and epidemic dengue 2 virus isolated in West Africa. *Trans R Soc Trop Med Hyg* **102**, 493–498.
- Diamond, M. S., Edgil, D., Roberts, T. G., Lu, B. & Harris, E. (2000). Infection of human cells by dengue virus is modulated by different cell types and viral strains. *J Virol* **74**, 7814–7823.
- 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.
- Eram, S., Setyabudi, Y., Sadono, T. I., Sutrisno, D. S., Gubler, D. J. & Sulianti Saroso, J. (1979). Epidemic dengue hemorrhagic fever in rural Indonesia. II. Clinical studies. *Am J Trop Med Hyg* **28**, 711–716.
- Fulmali, P. V., Walimbe, A. & Mahadev, P. V. (2008). Spread, establishment and prevalence of dengue vector *Aedes aegypti* (L.) in Konkan region, Maharashtra, India. *Indian J Med Res* **127**, 589–601.
- Gratz, N. G. (2004). Critical review of the vector status of *Aedes albopictus*. *Med Vet Entomol* **18**, 215–227.
- Gritsun, T. S., Holmes, E. C. & Gould, E. A. (1995). Analysis of flavivirus envelope proteins reveals variable domains that reflects their antigenicity and may determine their pathogenesis. *Virus Res* **35**, 307–321.
- Hayes, C. G., Phillips, I. A., Callahan, J. D., Griebenow, W. F., Hyams, K. C., Wu, S. J. & Watts, D. M. (1996). The epidemiology of dengue

- virus infection among urban, jungle, and rural populations in the Amazon region of Peru. *Am J Trop Med Hyg* 55, 459–463.
- Holmes, E. C. & Burch, S. S. (2000). The causes and consequences of genetic variation in dengue virus. *Trends Microbiol* 8, 74–77.
- Holmes, E. C. & Twiddy, S. S. (2003). The origin, emergence and evolutionary genetics of dengue virus. *Infect Genet Evol* 3, 19–28.
- Innis, B. L., Thriawuth, V. & Hemachudha, C. (1989). Identification of continuous epitopes of envelope glycoprotein of dengue type 2 virus. *Am J Trop Med Hyg* 40, 676–687.
- Kumar, A., Sharma, S. K., Padbidri, V. S., Thakare, J. P., Jain, D. C. & Datta, K. K. (2001). An outbreak of dengue fever in rural areas of northern India. *J Commun Dis* 33, 274–281.
- Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5, 150–163.
- Leitmeyer, K. C., Vaughn, D. W., Watts, D. M., Salas, R., De Chacon, I. V., Ramos, C. & Rico-Hesse, R. (1999). Dengue virus structural differences that correlate with pathogenesis. *J Virol* 73, 4738–4747.
- Mackerras, M. (1946). Transmission of dengue fever by *Aedes (Stegomyia) scutellaris* walk in New Guinea. *Trans R Soc Trop Med Hyg* 40, 295–312.
- Mahadev, P. V., Kollali, V. V., Rawal, M. L., Pujara, P. K., Shaikh, B. H., Ilkal, M. A., Pathak, V., Dhanda, V., Rodrigues, F. M. & Banerjee, K. (1993). Dengue in Gujarat state, India during 1988 and 1989. *Indian J Med Res* 97, 135–144.
- Mehendale, S. M., Risbud, A. R., Rao, J. A. & Banerjee, K. (1991). Outbreak of dengue fever in rural areas of Parbhani district of Maharashtra, India. *Indian J Med Res* 93, 6–11.
- Myat Thu, H., Lowry, K., Jiang, L., Hlaing, T., Holmes, E. C. & Aaskov, J. (2005). Lineage extinction and replacement in dengue type 1 virus populations are due to stochastic events rather than to natural selection. *Virology* 336, 163–172.
- Parida, M. M., Dash, P. K., Upadhyay, C., Saxena, P. & Jana, A. M. (2002). Serological and virological investigation of an outbreak of dengue fever in Gwalior, India. *Indian J Med Res* 116, 248–254.
- Pond, S. L. & Frost, S. D. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21, 2531–2533.
- Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Rambaut, A. & Drummond, A. (2007). Tracer version 1.4. University of Oxford, UK. <http://evolve.zoo.ox.ac.uk>
- Rey, F. A., Heinz, F. X., Mandl, C., Kunz, C. & Harrison, S. C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* 375, 291–298.
- Rico-Hesse, R. (1990). Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. *Virology* 174, 479–493.
- Rico-Hesse, R., Harrison, L. M., Salas, R. A., Tovar, D., Nisalak, A., Ramos, C., Boshell, J., de Mesa, M. T., Nogueira, R. M. & da Rosa, A. T. (1997). Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. *Virology* 230, 244–251.
- Roehrig, J. T., Risi, P. A., Brubaker, J. R., Hunt, A. R., Beaty, B. J., Trent, D. W. & Mathews, J. H. (1994). T-helper cell epitopes on the E-glycoprotein of dengue 2 Jamaica virus. *Virology* 198, 31–38.
- Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.
- Rosen, L., Rozeboom, L. E., Sweet, B. H. & Sabin, A. B. (1954). The transmission of dengue by *Aedes polynesiensis* marks. *Am J Trop Med Hyg* 3, 878–882.
- Sabin, A. B. (1952). Research on dengue during World War II. *Am J Trop Med Hyg* 1, 30–50.
- Self, L. S. (1984). Vector introductions associated with disease outbreaks in the western Pacific Region. *Dengue Newslett* 10, 50–51.
- Singh, U. B. & Seth, P. (2001). Use of nucleotide sequencing of the genomic cDNA fragments of the capsid/premembrane junction region for molecular epidemiology of dengue type 2 viruses. *Southeast Asian J Trop Med Public Health* 32, 326–335.
- Singh, U. B., Maitra, A., Baroor, S., Rai, A., Pasha, S. T. & Seth, P. (1999). Partial nucleotide sequencing and molecular evolution of epidemic causing dengue 2 strains. *J Infect Dis* 180, 959–965.
- Sittisombut, N., Sistayanarain, A., Cardosa, M. J., Salminen, M., Damrongdachakul, S., Kalayanarooj, S., Rojanasuphot, S., Supawadee, J. & Maneekarn, N. (1997). Possible occurrence of a genetic bottleneck in dengue serotype 2 viruses between the 1980 and 1987 epidemic seasons in Bangkok, Thailand. *Am J Trop Med Hyg* 57, 100–108.
- Song, M., Wang, B., Liu, J. & Gratz, N. (2003). Insect vectors and rodents arriving in China aboard international transport. *J Travel Med* 10, 241–244.
- Suchard, M. A., Weiss, R. E. & Sinsheimer, J. S. (2001). Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol* 18, 1001–1013.
- Sulianti, S. J. (1978). Dengue hemorrhagic fever in Indonesia: an epidemiological review. *Asian J Infect Dis* 2, 7–9.
- Swofford, D. L. (2003). PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4. Sunderland, MA: Sinauer Associates.
- Tewari, S. C., Thenmozhi, V., Katholi, C. R., Manavalan, R., Munirathinam, A. & Gajanana, A. (2004). Dengue vector prevalence and virus infection in a rural area in south India. *Trop Med Int Health* 9, 499–507.
- Twiddy, S. S., Farrar, J. J., Vinh, C. N., Wills, B., Gould, E., Gustin, T., Lloyd, G. & Holmes, E. C. (2002a). Phylogenetic relationships and differential selection pressures among genotypes of dengue 2 virus. *Virology* 298, 63–72.
- Twiddy, S. S., Woelk, C. H. & Holmes, E. C. (2002b). Phylogenetic evidence for adaptive evolution of dengue viruses in nature. *J Gen Virol* 83, 1679–1689.
- 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.
- Vasilakis, N., Holmes, E. C., Fokam, E. B., Faye, O., Diallo, M., Sall, A. A. & Weaver, S. C. (2007). Evolutionary processes among sylvatic dengue-2 viruses. *J Virol* 81, 9591–9595.
- Vasilakis, N., Deardorff, E. R., Kenney, J. L., Rossi, S. L., Hanley, K. A. & Weaver, S. C. (2009). Mosquitoes put the brake on arbovirus evolution: experimental evolution reveals slower mutation accumulation in mosquito than vertebrate cells. *PLoS Pathog* 5, e1000467. doi:10.1371/journal.ppat.1000467
- Wang, E., Ni, H., Xu, R., Barrett, A. D., Watowich, S. J., Gubler, D. J. & Weaver, S. C. (2000). Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. *J Virol* 74, 3227–3234.
- WHO (2002). *Fact Sheet: Dengue and Dengue Haemorrhagic Fever*. Geneva: World Health Organization.
- Wittke, V., Robb, T. E., Thu, H. M., Nimmannitya, S., Kalayanarooj, S., Vaughn, D. W., Endy, T. P., Holmes, E. C. & Aaskov, J. G. (2002). Extinction and rapid emergence of strains of dengue 3 virus during an interepidemic period. *Virology* 301, 148–156.
- Xu, G., Dong, H., Shi, N., Liu, S., Zhou, A., Cheng, Z., Chen, G., Liu, J., Fang, T. & other authors (2007). An outbreak of dengue virus serotype 1 infection in Cixi, Ningbo, People's Republic of China, 2004, associated with a traveler from Thailand and high density of *Aedes albopictus*. *Am J Trop Med Hyg* 76, 1182–1188.

Zanotto, P. M., Gould, E. A., Gao, G. F., Harvey, P. H. & Holmes, E. C. (1996). Population dynamics of flaviviruses revealed by molecular phylogenies. *Proc Natl Acad Sci U S A* **93**, 548–553.

Zeller, H. G. (1998). Dengue, arbovirus and migrations in the Indian Ocean. *Bull Soc Pathol Exot* **91**, 56–60.

Zhang, C., Mammen, M. P., Jr, Chinnawirotpisan, P., Klungthong, C., Rodpradit, P., Monkongdee, P., Nimmannitya, S., Kalayanaroj, S. & Holmes, E. C. (2005). Clade replacements in dengue virus serotypes 1 and 3 are associated with changing serotype prevalence. *J Virol* **79**, 15123–15130.