Journal of General Virology (2006), **87**, 873–883 **DOI** 10.1099/vir.0.81486-0

Structure and age of genetic diversity of dengue virus type 2 in Thailand

Chunlin Zhang,[1] Mammen P. Mammen, Jr,[1] Piyawan Chinnawirotpisan,[1] Chonticha Klungthong,[1] Prinyada Rodpradit,[1] Ananda Nisalak,[1] David W. Vaughn,[2] Suchitra Nimmannitya,[3] Siripen Kalayanarooj[3] and Edward C. Holmes[4]

Correspondence
Edward C. Holmes
ech15@psu.edu

[1]Department of Virology, US Army Medical Component-Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand

[2]Military Infectious Diseases Research Program, US Army Medical Research and Materiel Command, Fort Detrick, MD 21701, USA

[3]Queen Sirikit National Institute of Child Health, Bangkok, Thailand

[4]Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, PA 16802, USA

Dengue virus type 2 (DENV-2) is a common viral infection and an important health concern in South-East Asia. To determine the molecular evolution of DENV-2 in Thailand, 105 isolates of the E (envelope) gene and 10 complete genomes sampled over a 27 year period were sequenced. Phylogenetic analysis of these data revealed that three genotypes of DENV-2 have circulated in Thailand, although, since 1991, only viruses assigned to Asian genotype I have been sampled from the population. A broader analysis of 35 complete genomes of DENV-2 revealed that most amino acids are subject to strong selective constraints, indicative of widespread purifying selection against deleterious mutations. This was further supported by an analysis of genome-wide substitution rates, which indicated that DENV-2 fixes approximately 10 mutations per genome per year, far lower than expected from its mutational dynamics. Finally, estimates of the age of DENV-2 were remarkably consistent among genes, indicating that the current genetic diversity in this virus probably arose within the last 120 years, concordant with the first determination of the aetiology of dengue disease.

Received 1 September 2005
Accepted 1 December 2005

INTRODUCTION

Documenting the structure of genetic diversity in viral populations is central to understanding the factors responsible for their emergence. This task is greatly assisted by the phylogenetic analysis of viral gene sequence data, which allows the determination of various aspects of epidemiological history, including patterns of geographical movement, modes of population growth and rates of nucleotide substitution (Grenfell *et al.*, 2004). One virus where such analysis has been particularly informative is *Dengue virus* (DENV). In this case, phylogenetic analyses have provided valuable preliminary insights into the evolution of virulence and transmissibility that may then be tested by using experimental approaches.

DENV is a single-stranded, positive-sense RNA virus of the genus *Flavivirus* that comprises four serotypes (DENV-1 to DENV-4) and is transmitted among humans and other higher primates by mosquitoes of the genus *Aedes*. The importance of DENV lies in its high prevalence, with at least 50 million new cases each year (WHO, 2002), its association with some serious clinical conditions, particularly dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS), and because it is likely to increase in prevalence following urbanization, global travel and climate change.

Early phylogenetic studies of DENV uncovered abundant genetic variation within each serotype. Further, because this variation was organized as discrete clusters on trees, it could be classified into a series of genotypes or subtypes (Lanciotti *et al.*, 1994, 1997; Lewis *et al.*, 1993; Rico-Hesse, 1990). Most attention has been paid to DENV-2, within which six genotypes have been proposed, five of which are present in humans with differing geographical distributions (Twiddy *et al.*, 2002a). Two genotypes, 'Asian I' and 'Asian II', are currently restricted to South-East Asia, an 'American'

The GenBank/EMBL/DDBJ accession numbers for the DENV-2 sequences obtained in this study are DQ181807–DQ181901 for the E gene and DQ181797–DQ181806 for complete genomes.

Supplementary material is available in JGV Online.

genotype is now only found in the Americas and with decreasing frequency, an 'Asian/American' genotype has its ancestry in South-East Asia, but spread to the Americas in the early 1980s (Rico-Hesse *et al.*, 1997), and, finally, a 'Cosmopolitan' genotype has a wide distribution across the tropical and subtropical world (Twiddy *et al.*, 2002a). There is also growing experimental evidence that Asian strains of DENV-2, which are associated with DHF/DSS, are able to outcompete those sampled from the American genotype (Armstrong & Rico-Hesse, 2001; Cologna & Rico-Hesse, 2003; Cologna *et al.*, 2005), which are rarely associated with severe disease (Watts *et al.*, 1999).

More detailed molecular epidemiological studies have considered the spread of DENV-2 within specific localities, most notably the Caribbean (Foster *et al.*, 2004; Rodriguez-Roche *et al.*, 2005; Uzcategui *et al.*, 2001) and South-East Asia (Rico-Hesse *et al.*, 1998; Sittisombut *et al.*, 1997; Thu *et al.*, 2004; Trent *et al.*, 1989; Twiddy *et al.*, 2002a; Wittke *et al.*, 2002). One locality where such studies have been especially informative is Thailand. Not only is DENV a common paediatric disease in this country, but a series of studies has revealed both its epidemiological and evolutionary dynamics (Burke *et al.*, 1988; Cummings *et al.*, 2004; Endy *et al.*, 2004; Nisalak *et al.*, 2003; Rico-Hesse *et al.*, 1998). In particular, previous phylogenetic studies of DENV-2 in Thailand identified sporadic losses in genetic diversity, compatible with the action of population bottlenecks (Sittisombut *et al.*, 1997). However, whether these bottlenecks are due to the selective replacement of one strain by another, perhaps mediated by complex patterns of cross-immunity, or are caused by stochastic processes is unclear.

Herein, we document the evolution of DENV-2 in Thailand over a 27 year sampling period. In particular, we wished to determine whether there were major changes in the genotype composition of the DENV-2 strains circulating through time, the underlying rate of nucleotide substitution and selection pressures in the virus, and, from this estimate, the age of the current genetic diversity in DENV-2.

METHODS

Specimen data. DENV-2 isolates were obtained from C6/36 cell supernatant (specimens collected before 2000) or serum (specimens collected during and after the year 2000) kept at −70 °C at the Armed Forces Research Institute of Medical Sciences (AFRIMS). These were derived from 105 children (49 female, 56 male; mean age, 7·3 years) hospitalized at either the Queen Sirikit National Institute of Child Health (QSNICH) or at the Kamphaeng Phet Provincial Hospital (KPPPH) (northern Thailand) during the period 1974–2001 (Table 1). Grading of dengue disease for these specimens was conducted by using World Health Organization classification guidelines (WHO, 2002).

Identification of virus-specific serotype. All samples were identified as DENV-2 by using an antigen-capture ELISA as described by Henchal *et al.* (1983) and Kuno *et al.* (1985). The DENV-2 identification of all samples was also confirmed by using RT-PCR as described by Lanciotti *et al.* (1992). Primary versus secondary DENV-2 infection was determined solely by haemagglutination assay inhibition (HAI) for specimens collected before 1990 and by IgM/IgG ELISA supported by HAI for those specimens collected during and after 1990. HAI was performed by using the method of Clarke & Casals (1958). The anti-dengue IgM/IgG ELISA method used was described previously (Innis *et al.*, 1989).

RNA extraction and RT-PCR to generate DNA fragments used in sequencing. Virus RNA was extracted from cell-culture supernatant or serum by using a QIAamp Viral RNA Mini kit (Qiagen) according to the manufacturer's instructions. Oligonucleotide primers for generating eight overlapping DNA fragments (nt 1–1264, 798–2516, 2410–4093, 3953–5727, 5618–7562, 7426–9011, 8888–10057 and 9958–10723) spanning the entire DENV-2 genome sequence and oligonucleotide primers for sequencing were designed based on the consensus sequence obtained from 10 DENV-2 strains available in GenBank (accession numbers AF022437, AF119661, AF169678, AF169681, AF169684, AF169687, AF489932, AF208496, AF276619 and M20558) by using the Primer3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Genomic RNA was converted to cDNA by using random-hexamer oligonucleotides with the SuperScript First-Strand Synthesis system (Invitrogen) according to the manufacturer's instructions. All DNA fragments, including the envelope (E) gene and the other overlapping DNA fragments that covered the entire genome, were amplified by PCR with *Taq* DNA polymerase (Roche). The PCR-amplified DNA fragments were purified by using QIAquick PCR Purification kits and a QIAquick Gel Extraction kit (Qiagen) according to the manufacturer's instructions. Purified DNA fragments were used for sequencing.

Sequencing and analysis. Sequences of DENV-2 sequencing primers used are available in the Supplementary Table in JGV Online. Cycle-sequencing reactions were performed by using a DYEnamic ET Dye Terminator sequencing kit (Amersham Biosciences) according to the manufacturer's instructions. The sequencing products were cleaned by standard precipitation before sequencing in a MegaBACE 500 automated DNA sequencer (Amersham Biosciences). Overlapping nucleic acid sequences were combined for analysis and edited by using the SEQUENCHER software (Gene Codes Corporation).

Phylogenetic analysis. Three datasets were compiled for phylogenetic analysis (with identical sequences removed in all cases). First, to place the evolution of DENV-2 in Thailand in a global context, we compiled a dataset of 120 E gene sequences (1485 bp in length), representing the full extent of genetic diversity in DENV-2 and utilizing the four sylvatic strains as outgroups to root the tree. This dataset contained a representative sample of 36 of the Thai isolates sequenced here. Second, to reconstruct the evolutionary history of DENV-2 in Bangkok in particular, we inferred a phylogenetic tree for 79 E gene sequences of the viruses newly isolated during 1974–2001 from this city, considered the epicentre of dengue virus transmission within Thailand (Cummings *et al.*, 2004). This tree was rooted between the strains assigned to the Asian I and Asian/American genotypes. Finally, to understand the evolution of the individual genes of DENV-2, we analysed a representative sample of 35 complete genomes (coding regions only; 10 179 bp), including the sequences from 10 Thai isolates newly generated here. As no complete genomes from sylvatic strains are available, this tree was rooted by using a single DENV-1 isolate (Argentina.297/00), which was removed from the subsequent analysis. Trees were run individually for each of the 10 proteins and for the entire polyprotein.

For all datasets, maximum-likelihood (ML) phylogenetic trees were estimated by using PAUP* (Swofford, 2003). In all cases, we used the $GTR+I+\Gamma_4$ model of nucleotide substitution with successive rounds of branch swapping. To determine the support for individual groupings on the phylogenetic trees, we performed a bootstrap-resampling analysis using 1000 replicate neighbour-joining trees estimated under

the ML substitution model. To determine the position of the root in the complete-genome analysis, an ML tree was inferred by using a 3393 aa alignment under the WAG+Γ model available in the TREE-PUZZLE program (Strimmer & von Haeseler, 1996). Translated amino acids were used in this case because of the large genetic distance between DENV-2 and the DENV-1 outgroup sequence.

Measurement of selection pressures. For the 35 complete genomes of DENV-2, we measured the extent of genetic diversity and the underlying selection pressures by using the Datamonkey facility (Kosakovsky Pond & Frost, 2005). Genetic diversity was quantified as the total length of the tree for each gene in terms of the numbers of substitutions per site (denoted T_L). Overall and site-specific selection pressures were determined as the ratio of non-synonymous (d_N) to synonymous (d_S) substitutions per site, estimated by using the SLAC method (incorporating the HKY85 model of nucleotide substitution and phylogenetic trees inferred by using the neighbour-joining method). To obtain a less conservative analysis of selection pressures at individual sites, we also employed the CODEML program (models M7 and M8; Yang *et al.*, 2000), which uses an ML approach to fit the observed data to models of codon substitution that differ in their distribution of d_N/d_S ratios.

Rates and dates of molecular evolution. Two different datasets were used to estimate rates of nucleotide substitution in DENV-2, as well as the age of the most recent common ancestor (MRCA) of the viruses sampled. First, we examined the E gene sequences of the 79 Bangkok isolates, as sampling is relatively dense in this case. Second, to obtain rates and dates for the individual genes of DENV-2, we conducted an equivalent analysis for each gene of our complete-genome dataset of 35 viruses. All analyses were undertaken by using the Bayesian Markov Chain Monte Carlo (MCMC) method provided in the BEAST package (http://evolve.zoo.ox.ac.uk/beast/). This analysis was performed by using the GTR+I+Γ_4 substitution model under a coalescent model of constant population size, as this gave a better fit to the data under Akaike's information criterion than an exponential-growth model in all cases other than the C gene (see Results). In all cases, we employed a burn-in of 1 million and a final chain length of 10 million. To reveal the uncertainty in the estimation process, we also determined the 95 % high probability density (HPD) intervals in each case.

RESULTS AND DISCUSSION

Molecular epidemiology of DENV-2 in Thailand

Our phylogenetic analysis of 120 E genes representing the global diversity of DENV-2 confirmed the existence of six genotypes of this virus, five of which are associated with humans (Fig. 1). However, the evolutionary relationships among the genotypes are harder to determine, with none of the inter-genotype nodes supported by >70 % bootstrap replicates. Those strains sampled from Thailand fell into three of the human genotypes. The vast majority were identified as the Asian I genotype. However, five isolates sampled from 1980 to 1991 fell into the Asian/American genotype and a single Thai strain, D2.P8-377/69, sampled in 1969 fell into the Cosmopolitan genotype. Consequently, since 1991, all DENV-2 isolates sampled from Thailand belong to the Asian I genotype. Given the success of the Asian/American and Cosmopolitan genotypes in establishing DENV infections in other localities, it seems likely that the presumed disappearance of these viruses from Thailand is due to complex immunological interactions with other serotypes rather than any intrinsic fitness deficit compared with the Asian I viruses, although this will require more study.

Another noteworthy observation was the absence of strong spatial structure in the Thai data, as those viruses sampled from Kamphaeng Phet, some 358 km north of Bangkok, clustered closely with the Bangkok strains. In contrast, there was a clear clustering by time of sampling (see below). This pattern indicates that there is relatively free movement of viruses across Thailand during periods of high DENV transmission.

Our phylogenetic analysis of the 79 DENV-2 E gene sequences sampled from Bangkok similarly revealed a strong temporal structure, with the oldest sampled viruses tending to fall closest to the root of the tree (Fig. 2). This allowed us to undertake a relatively precise estimate of the rate of nucleotide substitution in DENV-2 by using Bayesian methods. This resulted in a rate of $8{\cdot}533 \times 10^{-4}$ substitutions per site per year (HPD, $7{\cdot}174 \times 10^{-4}$–$9{\cdot}952 \times 10^{-4}$ substitutions per site per year) and is hence within the range of most other estimates obtained for DENV (summarized by Twiddy *et al.*, 2003). The overall age of this tree, corresponding to the divergence between the Asian I and Asian/American genotypes, was approximately 57 years (HPD, 48–67 years). Hence, although the Asian/American genotype did not invade Latin America until the early 1980s (Rico-Hesse *et al.*, 1997), it has a deeper evolutionary separation from those viruses assigned to Asian genotype I.

Finally, we asked whether there was an association between viral strain, as defined by the E gene, and the clinical outcome of DENV infection. For this purpose, we examined the distribution of the strains associated with the most severe forms of dengue disease (grades III and IV, which signify DSS) and their position on our 79 sequence tree. This analysis provided no evidence that dengue fever (DF) and varying grades of DHF were associated with different viral strains, as those isolates sampled from patients with severe dengue disease were distributed evenly across the tree. Such a lack of association between E gene variation and clinical grade has been noted previously (Klungthong *et al.*, 2004; Rico-Hesse *et al.*, 1998; Shurtleff *et al.*, 2001). Further, the fact that there is no apparent increase in the frequency of DENV-2 strains associated with severe disease indicates that high virulence, itself associated with high viraemia (Vaughn *et al.*, 2000), does not guarantee long-term evolutionary success.

Analysis of complete genomes of DENV-2

To determine whether the evolutionary relationships observed for the E gene applied to the whole DENV-2 genome, we conducted a further phylogenetic analysis using the sequences of 35 whole genomes (coding regions only). As in the case of the E gene phylogeny, this tree clearly distinguishes among the five human genotypes of DENV-2, all with strong bootstrap support (Fig. 3). However, in

Table 1. DENV-2 isolates sequenced in this study

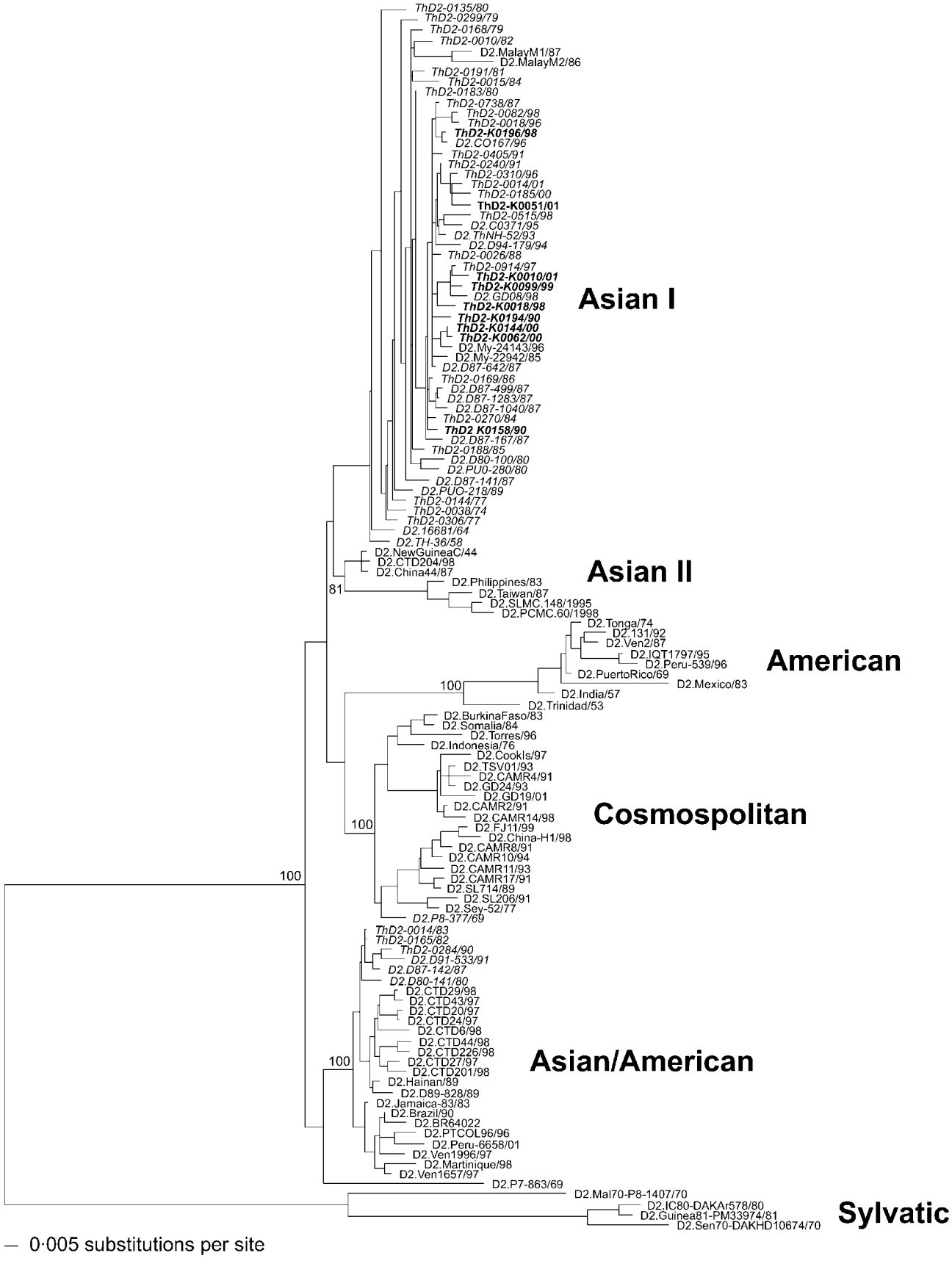
Samples shown in bold were chosen for complete-genome sequencing.

Virus sample*	Disease	Infection	Sex	Age (years)
ThD2-0038/74	DHF I	Secondary	F	8
ThD2-0066/74	DHF II	Secondary	M	3
ThD2-0100/76	DHF III	Secondary	F	12
ThD2-0161/76	DHF IV	Secondary	M	11
ThD2-0144/77	DHF III	Secondary	M	2
ThD2-0306/77	DHF II	Primary	F	0·41
ThD2-0051/78	DHF II	Primary	M	0·75
ThD2-0133/78	DHF II	Secondary	M	11
ThD2-0014/79	DF	Secondary	M	7
ThD2-0168/79	DHF IV	Secondary	F	8
ThD2-0299/79	DHF II	Primary	F	0·50
ThD2-0135/80	DF	Secondary	M	6
ThD2-0158/80	DF	Secondary	F	14
ThD2-0183/80	DHF IV	Secondary	F	2
ThD2-0184/80	DHF IV	Secondary	F	5
ThD2-0004/81	DF	Secondary	M	1·25
ThD2-0025/81	DHF II	Secondary	F	8
ThD2-0191/81	DHF IV	Primary	F	0·66
ThD2-0010/82	DHF II	Secondary	M	2
ThD2-0099/82	DHF III	Secondary	M	5
ThD2-0165/82	DHF I	Secondary	F	8
ThD2-0195/82	DHF III	Secondary	M	6
ThD2-0014/83	DHF IV	Secondary	M	5
ThD2-0185/83	DF	Secondary	M	10
ThD2-0307/83	DHF I	Primary	M	0·25
ThD2-0317/83	DHF III	Secondary	M	4
ThD2-0015/84	DF	Secondary	M	13
ThD2-0270/84	DHF IV	Secondary	M	6
ThD2-0498/84	DHF IV	Primary	M	14
ThD2-0501/84	DHF III	Primary	F	0·75
ThD2-0188/85	DF	Secondary	M	8
ThD2-0327/85	DHF IV	Secondary	M	6
ThD2-0433/85	DHF IV	Secondary	M	4
ThD2-0021/86	DF	Secondary	F	4
ThD2-0169/86	DHF III	Secondary	F	6
ThD2-0738/87	DHF IV	Secondary	M	9
ThD2-1121/87	DF	Secondary	M	13
ThD2-0026/88	DF	Secondary	M	13
ThD2-0032/88	DHF I	Secondary	M	4
ThD2-0372/88	DHF III	Secondary	M	12
ThD2-0284/90	DF	Secondary	M	11
ThD2-0847/90	DHF III	Primary	M	0·58
ThD2-1099/90	DHF III	Secondary	F	5
ThD2-1136/90	DHF III	Secondary	M	11
ThD2-0104/91	DF	Secondary	M	11
ThD2-0240/91	DHF III	Secondary	M	12
ThD2-0405/91	DHF III	Secondary	M	7
ThD2-0562/91	DHF III	Secondary	M	5
ThD2-0021/92	DHF II	Secondary	F	12
ThD2-0263/92	DHF II	Secondary	F	2
ThD2-0054/93	DF	Secondary	F	7·42
ThD2-0486/93	DHF IV	Secondary	M	6·42
ThD2-0562/93	DHF II	Secondary	M	8·83
ThD2-0587/93	DHF III	Secondary	M	4·66

Table 1. cont.

Virus sample*	Disease	Infection	Sex	Age (years)
ThD2-0035/94	DHF III	Secondary	M	4
ThD2-0082/94	DHF II	Secondary	F	7
ThD2-0249/94	DF	Primary	F	1·08
ThD2-0278/94	DHF IV	Secondary	F	12
ThD2-0194/95	DHF II	Primary	M	0·75
ThD2-0263/95	DF	Secondary	F	5
ThD2-0313/95	DHF III	Secondary	F	17
ThD2-0577/95	DHF I	Secondary	M	14
ThD2-0007/96	DHF III	Secondary	M	3
ThD2-0018/96	DHF II	Secondary	F	4
ThD2-0310/96	DF	Secondary	M	12
ThD2-0546/96	DHF IV	Secondary	F	11
ThD2-0009/97	DF	Secondary	M	5
ThD2-0914/97	DHFIV	Secondary	M	9
ThD2-1006/97	DHF III	Secondary	F	9
ThD2-0017/98	DF	Secondary	M	14
ThD2-0082/98	DHF III	Secondary	F	7
ThD2-0104/98	DHF II	Secondary	M	11
ThD2-0515/98	DHF IV	Secondary	M	9
ThD2-0055/99	DF	Secondary	F	6
ThD2-0568/99	DHF IV	Secondary	F	10
ThD2-0996/99	DHF III	Secondary	F	8
ThD2-1021/99	DHF II	Secondary	F	10
ThD2-0077/00	DHF II	Secondary	F	7
ThD2-0185/00	DF	Secondary	F	5
ThD2-0981/00	DHF IV	Secondary	F	6
ThD2-0014/01	DF	Secondary	F	8
ThD2-0078/01	DHF III	Secondary	F	8
ThD2-K0158/90†	DENV‡	Secondary	F	10
ThD2-K0194/90†	DENV‡	Secondary	F	10
ThD2-K0304/93†	DENV‡	Secondary	M	9
ThD2-K0307/93†	DENV‡	Secondary	M	7
ThD2-K0390/93†	DENV‡	Secondary	M	12
ThD2-K0013/94†	DENV‡	Secondary	M	8
ThD2-K0031/94†	DENV‡	NA	F	5
ThD2-K0001/95†	DENV‡	Secondary	F	12
ThD2-K0031/95†	DENV‡	Secondary	F	7
ThD2-K0003/96†	DENV‡	Secondary	M	8
ThD2-K0035/96†	DENV‡	Secondary	M	7
ThD2-K0123/97†	DENV‡	Secondary	F	5
ThD2-K0344/97†	DENV‡	Secondary	F	10
ThD2-K0018/98†	DENV‡	Secondary	M	8
ThD2-K0196/98†	DENV‡	Secondary	F	2
ThD2-K0099/99†	DENV‡	Secondary	F	4
ThD2-K0185/99†	DENV‡	Secondary	M	11
ThD2-K0012/00†	DENV‡	Secondary	M	2
ThD2-K0062/00†	DENV‡	Secondary	M	16
ThD2-K0123/00†	DENV‡	Secondary	F	11
ThD2-K0144/00†	DENV‡	Secondary	F	5
ThD2-K0010/01†	DENV‡	Secondary	F	9
ThD2-K0057/01†	DENV‡	Secondary	F	11

*The middle four numerals are the sample number, whilst the last two numerals indicate the year of isolation.
†These samples were collected at KPPPH in northern Thailand.
‡These samples were not clinically graded.


ThD2-0135/80
ThD2-0299/79
ThD2-0168/79
ThD2-0010/82
D2.MalayM1/87
D2.MalayM2/86
ThD2-0191/81
ThD2-0015/84
ThD2-0183/80
ThD2-0738/87
ThD2-0082/98
ThD2-0018/96
ThD2-K0196/98
D2.CO167/96
ThD2-0405/91
ThD2-0240/91
ThD2-0310/96
ThD2-0014/01
ThD2-0185/00
ThD2-K0051/01
ThD2-0515/98
D2.C0371/95
D2.ThNH-52/93
D2.D94-179/94
ThD2-0026/88
ThD2-0914/97
ThD2-K0010/01
ThD2-K0099/99
D2.GD08/98
ThD2-K0018/98
ThD2-K0194/90
ThD2-K0144/00
ThD2-K0062/00
D2.My-24143/96
D2.My-22942/85
D2.D87-642/87
ThD2-0169/86
D2.D87-499/87
D2.D87-1283/87
D2.D87-1040/87
ThD2-0270/84
ThD2 K0158/90
D2.D87-167/87
ThD2-0188/85
D2.D80-100/80
D2.PU0-280/80
D2.D87-141/87
D2.PUO-218/89
ThD2-0144/77
ThD2-0038/74
ThD2-0306/77
D2.16681/64
D2.TH-36/58
D2.NewGuineaC/44
D2.CTD204/98
D2.China44/87
D2.Philippines/83
D2.Taiwan/87
D2.SLMC.148/1995
D2.PCMC.60/1998
D2.Tonga/74
D2.131/92
D2.Ven2/87
D2.IQT1797/95
D2.Peru-539/96
D2.PuertoRico/69
D2.Mexico/83
D2.India/57
D2.Trinidad/53
D2.BurkinaFaso/83
D2.Somalia/84
D2.Torres/96
D2.Indonesia/76
D2.CookIs/97
D2.TSV01/93
D2.CAMR4/91
D2.GD24/93
D2.GD19/01
D2.CAMR2/91
D2.CAMR14/98
D2.FJ11/99
D2.China-H1/98
D2.CAMR8/91
D2.CAMR10/94
D2.CAMR11/93
D2.CAMR17/91
D2.SL714/89
D2.SL206/91
D2.Sey-52/77
D2.P8-377/69
ThD2-0014/83
ThD2-0165/82
ThD2-0284/90
D2.D91-533/91
D2.D87-142/87
D2.D80-141/80
D2.CTD29/98
D2.CTD43/97
D2.CTD20/97
D2.CTD24/97
D2.CTD6/98
D2.CTD44/98
D2.CTD226/98
D2.CTD27/97
D2.CTD201/98
D2.Hainan/89
D2.D89-828/89
D2.Jamaica-83/83
D2.Brazil/90
D2.BR64022
D2.PTCOL96/96
D2.Peru-6658/01
D2.Ven1996/97
D2.Martinique/98
D2.Ven1657/97
D2.P7-863/69
D2.Mal70-P8-1407/70
D2.IC80-DAKAr578/80
D2.Guinea81-PM33974/81
D2.Sen70-DAKHD10674/70
Asian I
Asian II
American
Cosmospolitan
Asian/American
Sylvatic
81
100
100
100
100
0·005 substitutions per site

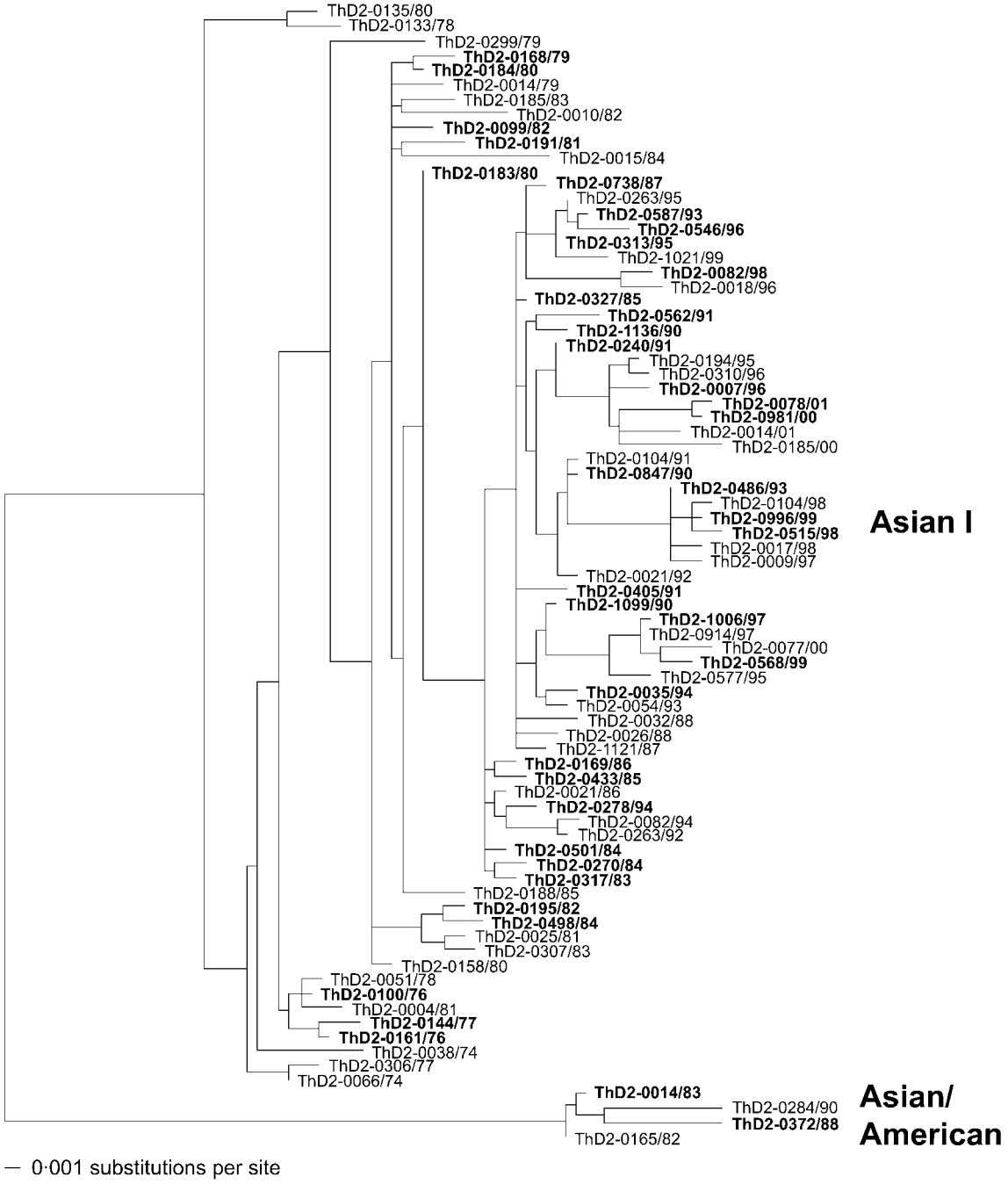


Fig. 2. ML phylogenetic tree of 79 E gene sequences representing the genetic diversity of DENV-2 in Bangkok, Thailand, during the period 1974–2001. Those isolates associated with grades III and IV of DHF/DSS, the most serious manifestations of dengue disease, are shown in bold.

contrast to the E gene tree, those viruses assigned to the American genotype are the most divergent. Such a change in phylogenetic position could reflect the divergent nature of the outgroup sequence used in this case, as the mean amino acid distance between the DENV-1 and DENV-2 isolates is approximately 40 % under the ML substitution model.

Fig. 1. ML phylogenetic tree of 120 E gene sequences representing the global diversity of DENV-2. The genotypes of DENV-2 are indicated, as are their associated bootstrap support values. All isolates sampled from Thailand are shown in italics, with those sampled from KPPPH shown in bold italics. The tree is rooted between the human and sylvatic strains of DENV-2 and all horizontal branch lengths are drawn to a scale of substitutions per site.

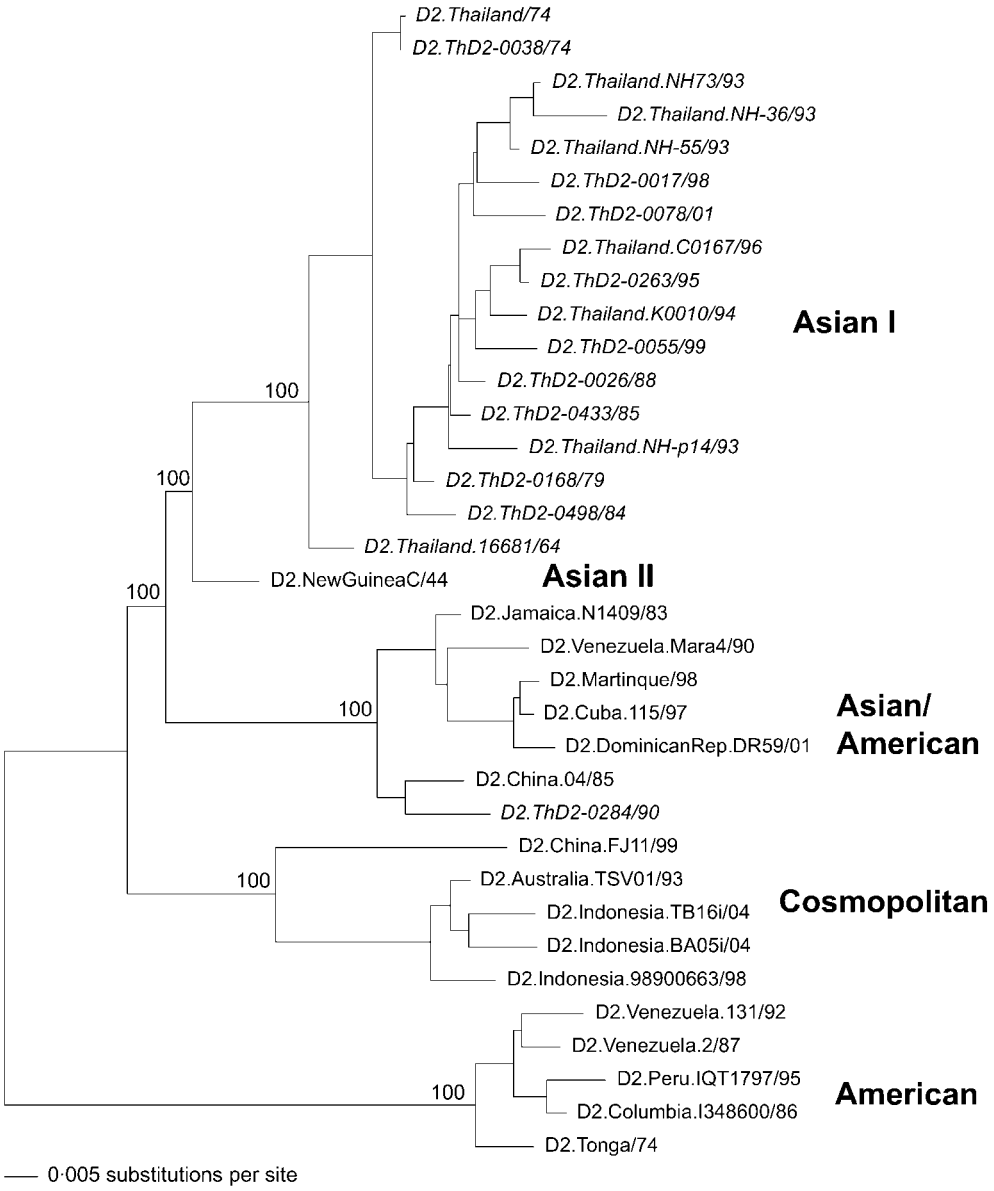


Fig. 3. ML phylogenetic tree of 35 complete coding-region sequences of DENV-2. The genotypes of DENV-2 are indicated, as are their associated bootstrap support values, and all isolates sampled from Thailand are shown in italics. The tree is rooted between the American and the other genotypes, as specified when a DENV-1 outgroup sequence was added to the dataset.

Another notable difference with the E gene tree is that the Asian/American genotype is now related more closely to the Asian I and II genotypes than to the Cosmopolitan genotype. Moreover, there is strong bootstrap support for this branching order, with all relevant nodes found in 100 % of replicates. However, individual gene trees varied greatly in topology, generating six different phylogenies (see Supplementary Figure, available in JGV Online). The phylogeny in four of the 10 genes (C, NS1, NS4B and NS5) was as that depicted in the whole-genome tree. In contrast, the Cosmopolitan and Asian/American genotypes formed the closest pairing in the E and NS4A genes and a variety of other patterns were seen in other genes. Given that the inter-genotypic nodes in these individual gene trees are often associated with low bootstrap values, such variable phylogenies seem most compatible with a lack of phylogenetic resolution caused by a rapid genetic diversification over the last century.

For the 35 complete genomes of DENV-2, we conducted a more precise analysis of evolutionary processes (Table 2). There was relatively little variation (1·6-fold) in the overall extent of genetic diversity among the individual genes of DENV-2 as measured by total tree length, although there was rather more in the overall rates of nucleotide substitution (2·8-fold, although most HPD values overlap). The tree

Table 2. Evolutionary processes among the proteins of DENV-2

Protein	Length (codons)	T_L*	d_N/d_S	Rate $\times 10^{-4}$† (HPD)	Age in years (HPD)
Capsid	114	0·358	0·099	3·956 (1·736–6·323)	174 (91–283)
Membrane	166	0·478	0·089	8·421 (6·040–10·790)	106 (80–135)
Envelope	495	0·436	0·050	8·250 (6·753–9·802)	108 (91–126)
NS1	352	0·472	0·070	7·851 (6·223–9·516)	114 (95–138)
NS2A	218	0·578	0·071	11·160 (8·715–13·600)	98 (80–118)
NS2B	130	0·454	0·049	8·588 (5·592–11·600)	130 (89–178)
NS3	618	0·491	0·033	8·778 (7·392–10·230)	107 (92–122)
NS4A	149	0·555	0·040	9·522 (6·653–12·380)	120 (88–156)
NS4B	249	0·495	0·037	8·184 (6·213–10·230)	119 (94–147)
NS5	900	0·483	0·054	6·906 (5·924–7·912)	124 (108–142)

*Tree length in substitutions per site.
†Rate of nucleotide substitution per site per year.

with the shortest length and, hence, the lowest rate of nucleotide substitution was from the capsid ($T_L = 0{\cdot}358$, mean rate $= 3{\cdot}956 \times 10^{-4}$ substitutions per site per year), whilst that with the longest length and thus the highest rate of evolutionary change was NS2A ($T_L = 0{\cdot}578$, mean rate $= 11{\cdot}160 \times 10^{-4}$ substitutions per site per year). Consequently, if the mean substitution rate across the DENV-2 genome is approximately 10^{-3} substitutions per site per year, then approximately 10 mutations are fixed per 10 kb genome per year. This rate is clearly lower than most estimates of the intrinsic mutation rate in RNA viruses, at roughly one mutation per genome per replication (Drake & Holland, 1999). The most likely explanation for such a disparity in rate is that the majority of mutations that arise during DENV replication are deleterious and removed by purifying selection, thereby reducing the long-term substitution rate.

The basis of the elevated substitution rate in NS2A is uncertain, particularly as the function of this protein remains unclear (Lindenbach & Rice, 2001). A recent study of DENV-4 in Thailand also identified NS2A as the most variable gene (Klungthong *et al.*, 2004) and an analysis of DENV-4 evolution in Puerto Rico identified likely positive selection at three amino acid sites in NS2A, although their functional consequences could not be ascertained (Bennett *et al.*, 2003). It is therefore evident that more work is required to identify the cause of the elevated levels of genetic variation in NS2A, and particularly whether it might be associated with escape from cytotoxic T-lymphocyte (CTL) recognition, which has been shown to be important in shaping evolutionary patterns in NS3 (Mongkolsapaya *et al.*, 2003; Simmons *et al.*, 2005).

In an attempt to determine their overall selective landscape, we compared the ratio of non-synonymous to synonymous substitutions in the individual genes of DENV-2. Again, there was relatively little variation (threefold) among genes, with the highest d_N/d_S value seen in the capsid gene (0·099) and lowest in NS3 (0·033), which encodes the viral helicase. Overall, such low d_N/d_S ratios are again indicative of the action of strong purifying selection, which appears to be true of arthropod-borne viruses as a whole (Woelk & Holmes, 2002). This was confirmed in a codon-specific analysis of selection pressures using both the SLAC and CODEML methods. Under the former, we found no evidence for positive selection in our complete-genome alignment, whereas the latter detected adaptive evolution ($P < 0{\cdot}05$) at only two codons, amino acid positions 135 and 637 in NS5, both of which have been identified previously (Twiddy *et al.*, 2002b; full results available from authors on request).

Finally, we explored the age of the sampled genetic diversity in human DENV-2. Most genes gave the age of the MRCA of human DENV-2 as approximately 100–120 years ago, with relatively tight HPD values (80–178 years). The exception was the capsid gene, which gave a mean age of the MRCA of 174 years ago and had relatively large HPD values (91–283 years). Moreover, whilst all other genes of DENV-2 supported a model of constant population size through time, the capsid gene favoured a model of exponential population growth. It is therefore clear that the capsid gene differs in its evolutionary dynamics from the other DENV-2 genes; it is evolving more slowly, has accumulated proportionally more amino acid substitutions and has older polymorphism. The underlying cause of these differing evolutionary dynamics is unclear, although it is known that the DENV capsid protein, which performs the critical role of RNA genome encapsidation, also contains CTL epitopes (Gagnon *et al.*, 1996), as well as potential B-cell epitopes (Anandarao *et al.*, 2005). However, the low substitution rate also hints at the existence of RNA secondary structures.

Overall, our estimates of the time of the MRCA of DENV-2 in humans are compatible with previous estimates (Twiddy *et al.*, 2003) and suggest that current diversity in DENV arose at the end of the 19th and beginning of the 20th centuries, when the aetiology of dengue disease was first described (Gubler, 2004). Such a recent estimate also means

that no DENV-2 lineages from the earliest described dengue epidemics, at the end of the 18th century (Gubler, 1997), have survived to the present day. The burgeoning genetic diversity of DENV, such as that demonstrated here in Thailand, is probably a function of the hyperendemicity of the virus and will also facilitate the production of strains with varying phenotypic properties, including virulence and transmissibility.

ACKNOWLEDGEMENTS

This research was supported by the US Military Infectious Diseases Research Program of the US Department of Defense, Fort Detrick, MD, USA. We thank the doctors and nursing staff of QSNICH for their outstanding patient care and illness grading, and the staff members of the Department of Virology, AFRIMS, who over the past 27 years have carefully performed diagnostic testing, data entry and specimen archival. The opinions and assertions contained herein are not to be construed as official or as reflecting the views of the US Government.

REFERENCES

Anandarao, R., Swaminathan, S. & Khanna, N. (2005). The identification of immunodominant linear epitopes of dengue type 2 virus capsid and NS4a proteins using pin-bound peptides. *Virus Res* **112**, 60–68.

Armstrong, P. M. & Rico-Hesse, R. (2001). Differential susceptibility of *Aedes aegypti* to infection by the American and Southeast Asian genotypes of dengue type 2 virus. *Vector Borne Zoonotic Dis* **1**, 159–168.

Bennett, S. N., Holmes, E. C., Chirivella, M., Rodriguez, D. M., Beltran, M., Vorndam, V., Gubler, D. J. & McMillan, W. O. (2003). Selection-driven evolution of emergent dengue virus. *Mol Biol Evol* **20**, 1650–1658.

Burke, D. S., Nisalak, A., Johnson, D. E. & Scott, R. M. (1988). A prospective study of dengue infections in Bangkok. *Am J Trop Med Hyg* **38**, 172–180.

Clarke, D. H. & Casals, J. (1958). Techniques for hemagglutination and hemagglutination-inhibition with arthropod-borne viruses. *Am J Trop Med Hyg* **7**, 561–573.

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.

Cummings, D. A. T., Irizarry, R. A., Huang, N. E., Endy, T. P., Nisalak, A., Ungchusak, K. & Burke, D. S. (2004). Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. *Nature* **427**, 344–347.

Drake, J. W. & Holland, J. J. (1999). Mutation rates among RNA viruses. *Proc Natl Acad Sci U S A* **96**, 13910–13913.

Endy, T. P., Nisalak, A., Chunsuttitwat, S., Vaughn, D. W., Green, S., Ennis, F. A., Rothman, A. L. & Libraty, D. H. (2004). Relationship of preexisting dengue virus (DV) neutralizing antibody levels to viremia and severity of disease in a prospective cohort study of DV infection in Thailand. *J Infect Dis* **189**, 990–1000.

Foster, J. E., Bennett, S. N., Carrington, C. V. F., Vaughan, H. & McMillan, W. O. (2004). Phylogeography and molecular evolution of dengue 2 in the Caribbean basin, 1981–2000. *Virology* **324**, 48–59.

Gagnon, S. J., Zeng, W., Kurane, I. & Ennis, F. A. (1996). Identification of two epitopes on the dengue 4 virus capsid protein recognized by a serotype-specific and a panel of serotype-cross-reactive human $CD4^+$ cytotoxic T-lymphocyte clones. *J Virol* **70**, 141–147.

Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A. & Holmes, E. C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–332.

Gubler, D. J. (1997). Dengue and dengue hemorrhagic fever: its history and resurgence as a global public health problem. In *Dengue and Dengue Hemorrhagic Fever*, pp. 1–22. Edited by D. J. Gubler & G. Kuno. London: CAB International.

Gubler, D. J. (2004). Commentary: Ashburn PM, Craig CF. Experimental investigations regarding the etiology of dengue. J Infect Dis 1907;4:440–75. *J Infect Dis* **189**, 1744–1783.

Henchal, E. A., McCown, J. M., Seguin, M. C., Gentry, M. K. & Brandt, W. E. (1983). Rapid identification of dengue virus isolates by using monoclonal antibodies in an indirect immunofluorescence assay. *Am J Trop Med Hyg* **32**, 164–169.

Innis, B. L., Nisalak, A., Nimmannitya, S., Kusalerdchariya, S., Chongswasdi, V., Suntayakorn, S., Puttisri, P. & Hoke, C. H. (1989). An enzyme-linked immunosorbent assay to characterize dengue infections where dengue and Japanese encephalitis co-circulate. *Am J Trop Med Hyg* **40**, 418–427.

Klungthong, C., Zhang, C., Mammen, M. P., Jr, Ubol, S. & Holmes, E. C. (2004). The molecular epidemiology of dengue virus serotype 4 in Bangkok, Thailand. *Virology* **329**, 168–179.

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.

Kuno, G., Gubler, D. J. & Santiago de Weil, N. S. (1985). Antigen capture ELISA for the identification of dengue viruses. *J Virol Methods* **12**, 93–103.

Lanciotti, R. S., Calisher, C. H., Gubler, D. J., Chang, G.-J. & Vorndam, A. V. (1992). Rapid detection and typing of dengue viruses from clinical samples by using reverse transcriptase-polymerase chain reaction. *J Clin Microbiol* **30**, 545–551.

Lanciotti, R. S., Lewis, J. G., Gubler, D. J. & Trent, D. W. (1994). Molecular evolution and epidemiology of dengue-3 viruses. *J Gen Virol* **75**, 65–75.

Lanciotti, R. S., Gubler, D. J. & Trent, D. W. (1997). Molecular evolution and phylogeny of dengue-4 viruses. *J Gen Virol* **78**, 2279–2286.

Lewis, J. A., Chang, G.-J., Lanciotti, R. S., Kinney, R. M., Mayer, L. W. & Trent, D. W. (1993). Phylogenetic relationships of dengue-2 viruses. *Virology* **197**, 216–224.

Lindenbach, B. D. & Rice, C. M. (2001). *Flaviviridae*: the viruses and their replication. In *Fields Virology*, 4th edn, pp. 991–1042. Edited by D. M. Knipe & P. M. Howley. Philadelphia: Lippincott, Williams & Wilkins.

Mongkolsapaya, J., Dejnirattisai, W., Xu, X.-N. & 11 other authors (2003). Original antigenic sin and apoptosis in the pathogenesis of dengue hemorrhagic fever. *Nat Med* **9**, 921–927.

Nisalak, A., Endy, T. P., Nimmannitya, S. & 7 other authors (2003). Serotype-specific dengue virus circulation and dengue disease in Bangkok, Thailand from 1973 to 1999. *Am J Trop Med Hyg* **68**, 191–202.

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. & 7 other authors (1997). Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. *Virology* **230**, 244–251.

Rico-Hesse, R., Harrison, L. M., Nisalak, A., Vaughn, D. W., Kalayanarooj, S., Green, S., Rothman, A. L. & Ennis, F. A. (1998).

Molecular evolution of dengue type 2 virus in Thailand. *Am J Trop Med Hyg* **58**, 96–101.

Rodriguez-Roche, R., Alvarez, M., Gritsun, T., Halstead, S., Kouri, G., Gould, E. A. & Guzman, M. G. (2005). Virus evolution during a severe dengue epidemic in Cuba, 1997. *Virology* **334**, 154–159.

Shurtleff, A. C., Beasley, D. W. C., Chen, J. J. Y. & 9 other authors (2001). Genetic variation in the 3′ non-coding region of dengue viruses. *Virology* **281**, 75–87.

Simmons, C. P., Dong, T., Chau, N. V. & 7 other authors (2005). Early T-cell responses to dengue virus epitopes in Vietnamese adults with secondary dengue virus infections. *J Virol* **79**, 5665–5675.

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.

Strimmer, K. & von Haeseler, A. (1996). Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol Biol Evol* **13**, 964–969.

Swofford, D. L. (2003). PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods), version 4. Sunderland, MA: Sinauer Associates.

Thu, H. M., Lowry, K., Myint, T. T., Shwe, T. N., Han, A. M., Khin, K. K., Thant, K. Z., Thein, S. & Aaskov, J. (2004). Myanmar dengue outbreak associated with displacement of serotypes 2, 3, and 4 by dengue 1. *Emerg Infect Dis* **10**, 593–597.

Trent, D. W., Grant, J. A., Monath, T. P., Manske, C. L., Corina, M. & Fox, G. E. (1989). Genetic variation and microevolution of dengue 2 virus in Southeast Asia. *Virology* **172**, 523–535.

Twiddy, S. S., Farrar, J. F., Chau, N. V., Wills, B., Gould, E. A., Gritsun, 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.

Uzcategui, N. Y., Camacho, D., Comach, G., Cuello de Uzcategui, R., Holmes, E. C. & Gould, E. A. (2001). Molecular epidemiology of dengue type 2 virus in Venezuela: evidence for *in situ* virus evolution and recombination. *J Gen Virol* **82**, 2945–2953.

Vaughn, D. W., Green, S., Kalayanarooj, S. & 8 other authors (2000). Dengue viremia titer, antibody response pattern, and virus serotype correlate with disease severity. *J Infect Dis* **181**, 2–9.

Watts, D. M., Porter, K. R., Putvatana, P., Vasquez, B., Calampa, C., Hayes, C. G. & Halstead, S. B. (1999). Failure of secondary infection with American genotype dengue 2 to cause dengue haemorrhagic fever. *Lancet* **354**, 1431–1434.

WHO (2002). Dengue and dengue haemorrhagic fever. Fact sheet no. 117. http://www.who.int/mediacentre/factsheets/fs117/en/

Wittke, V., Robb, T. E., Thu, H. M. & 7 other authors (2002). Extinction and rapid emergence of strains of dengue 3 virus during an interepidemic period. *Virology* **301**, 148–156.

Woelk, C. H. & Holmes, E. C. (2002). Reduced positive selection in vector-borne RNA viruses. *Mol Biol Evol* **19**, 2333–2336.

Yang, Z., Nielsen, R., Goldman, N. & Krabbe Pedersen, A.-M. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431–449.