Journal of General Virology (2002), **83**, 1679–1689. Printed in Great Britain

Phylogenetic evidence for adaptive evolution of dengue viruses in nature

S. Susanna Twiddy, Christopher H. Woelk and Edward C. Holmes

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

A maximum-likelihood approach was used to analyse selection pressures acting on genes from all four serotypes of dengue virus (DEN). A number of amino acid positions were identified within the envelope (E) glycoprotein that have been subject to relatively weak positive selection in both DEN-3 and DEN-4, as well as in two of the five genotypes of DEN-2. No positive selection was detected in DEN-1. In accordance with the function of the E protein as the major antigenic determinant of DEN, the majority of these sites were located in, or near to, potential T- or B-cell epitopes. A smaller number of selected sites was located in other well-defined functional domains of the E protein, suggesting that cell tropism and virus-mediated membrane fusion may also confer fitness advantages to DEN in nature. Several positively selected amino acid substitutions were also identified in the NS2B and NS5 genes of DEN-2, although the cause of this selection is unclear, whereas the capsid, membrane and non-structural genes NS1, NS2A, NS3 and NS4 were all subject to strong functional constraints. Hence, evidence was found for localized adaptive evolution in natural isolates of DEN, revealing that selection pressures differ among serotypes, genotypes and viral proteins.

Introduction

Dengue virus (DEN) is the agent of an important arbovirus disease, with an estimated annual infection rate in excess of 50 million (WHO, 2000), of which the majority are 'silent', with no overt clinical symptoms. However, a significant minority of infected individuals go on to develop life-threatening dengue haemorrhagic fever/dengue shock syndrome, which has an increasing incidence in tropical and subtropical countries.

DEN has a positive-sense, single-stranded RNA genome (genus *Flavivirus*), which exists as four genetically and antigenically distinct serotypes, denoted DEN-1 to -4. The viral genome is approximately 11 kb in length and consists of a non-translated region (NTR) of ~100 bp at the 5′ end followed by a single open reading frame encoding a polypeptide of approximately 3400 aa, post-translationally cleaved to produce three structural and seven non-structural proteins in the order C–prM/M–E–NS1–NS2A–NS2B–NS3–NS4A–NS4B–NS5. This is followed by a 3′ NTR of ~450 bp.

Recent studies have shown that strains of DEN may differ in their ability to infect cells and to cause disease (Leitmeyer *et al.*, 1999; Diamond *et al.*, 2000). Given that virus genetic diversity may therefore influence disease, it is clearly desirable to understand the evolutionary processes that generate genetic variation in natural populations of DEN. In particular, it is important to determine whether there is evidence of adaptive evolution in DEN, such as that driven by immune selection pressure, as well as the precise genomic regions involved. Despite the growing interest in the molecular epidemiology of DEN (Lewis *et al.*, 1993; Rico-Hesse *et al.*, 1998), there have been few attempts to determine which regions of the DEN genome are subject to positive selection, although this may be a key indicator of the nature of the interaction between host and virus. Indeed, it is generally considered that the most common pressure acting on DEN in nature is purifying selection, with little or no evidence of adaptive evolution produced to date (Zanotto *et al.*, 1996).

Herein, we present the results of a maximum-likelihood (ML) analysis of selection pressures acting on DEN, utilizing comparisons of the ratio of non-synonymous (d_N) to synonymous (d_S) substitutions (d_N/d_S, parameter ω) in all genes and serotypes. The ω ratio is a powerful indicator of the strength of natural selection acting on gene sequences,

Author for correspondence: Susanna Twiddy.
Fax +44 1865 310447. e-mail Susanna.Twiddy@zoo.ox.ac.uk

0001-8285

Table 1. DEN data sets used in this study

The APT data set contains all passage types, whereas the MPO data set contains isolates that were only passaged through mosquito cells.

Data set	Taxa	Length (codons)	Data set	Taxa	Length (codons)
DEN-1 E	9	495	DEN-2 E (Asian 2, VN)	9	495
DEN-2 E (APT)	53	495	DEN-2 E (Cosmopolitan)	29	495
DEN-2 E (MPO)	40	495	DEN-2 C	18	113
DEN-3 E (APT)	26	493	DEN-2 prM/M	19	166
DEN-3 E (MPO)	19	493	DEN-2 NS1	30	352
DEN-4 E (APT)	21	495	DEN-2 NS2A	25	218
DEN-4 E (MPO)	18	495	DEN-2 NS2B	27	130
DEN-2 E (American)	10	495	DEN-2 NS3	31	618
DEN-2 E (American/Asian)	43	495	DEN-2 NS4A	27	150
DEN-2 E (Asian 1)	23	495	DEN-2 NS4B	24	248
DEN-2 E (Asian 2)	19	495	DEN-2 NS5	36	900
DEN-2 E (Asian 2, PH)	11	495			

including those from RNA viruses (reviewed by Yang & Bielawski, 2000). Amino acid substitutions that are selectively neutral will be fixed at the same rate as neutral synonymous changes, so that $\omega = 1$, while the operation of negative (purifying) selection will result in a reduction in the rate at which non-synonymous mutations are fixed, giving an ω ratio < 1. In contrast, where an amino acid change increases virus fitness, it will be fixed at a higher rate than a synonymous mutation subject only to genetic drift, resulting in an ω ratio > 1. Most previous analyses of ω ratios have relied on multiple pairwise comparisons of each sequence in a data set. Although informative, such methods are hampered by a lack of independence, do not consider that individual codons may differ in selection pressure or are often based on unrealistic models of nucleotide substitution – for example, they may assume equal rates for transitions and transversions or uniform codon usage and hence may miss localized examples of positive selection (Zanotto *et al.*, 1999). Consequently, the new generation of analytical methods that analyse ω ratios codon-by-codon take into account the phylogenetic relationships of the sequences in question, utilize realistic models of nucleotide substitution and employ rigorous statistics, such as the likelihood ratio test (LRT), to compare ω ratios, representing a major advance (Yang *et al.*, 2000). It is these methods that we employ here.

Like most molecular studies of DEN and other flaviviruses, the main focus of our analysis is the envelope (E) gene, which encodes the major protein component of the virion surface, is the most important antigen with regard to humoral immunity (Henchal *et al.*, 1985; Mandl *et al.*, 1988, 1989; Innis *et al.*, 1989; Gritsun *et al.*, 1995) and is associated with other biological activities, including cell attachment/receptor binding and virus assembly. Our analysis considers the E gene of each serotype separately. For DEN-2, for which most sequence data are available, we also investigate the nature of selection pressures acting on other viral genes that have a variety of functional roles.

Methods

■ **Strains used and sequence analysis.** We compiled three major data sets of DEN gene sequences, from which subsets were used for each analysis (Table 1). The first comprised 109 DEN E genes collected from GenBank, representing all four virus serotypes. This included 35 new DEN-2 sequences sampled from a variety of locations, most notably Vietnam (Twiddy *et al.*, 2002). For serotypes other than DEN-1, for which a limited number of sequences was available, all strains were reclassified into data sets comprising all passage types (APT) and mosquito passage only (MPO), and these were analysed separately. This analysis was performed because it has been reported previously that passaging may introduce artificial evidence for positive selection (Woelk *et al.*, 2001). A full list of the sequences used in this analysis is available at JGV Online as supplementary data (http://vir.sgmjournals.org) or from the authors on request.

The second data set consisted of 171 E gene sequences from DEN-2 only. This data set comprised all DEN-2 sequences available in GenBank, excluding possible recombinants, as well as the 35 new sequences described above. A prior phylogenetic analysis of these data indicated that human DEN-2 viruses formed five genotypes, designated American, Asian 1 and Asian 2, American/Asian and Cosmopolitan (Fig. 1), and each was analysed separately. The Asian 2 genotype was further divided into Vietnamese (VN) and Philippino (PH) clades, each of which was well supported on the phylogenetic tree. These two subsets were again analysed separately.

The third data set consisted of all (36) available whole genome sequences for DEN-2. In this case, the set of sequences for each gene, excluding the E gene, was analysed separately.

For each data set, sequences were aligned using CLUSTALW (Thompson *et al.*, 1994) and checked manually. Identical sequences, vaccine strains, sequences containing frameshifts and putative recombinants, as well as sylvatic (non-human primate and sylvatic vector) strains, were removed in all cases. Putative recombinants were determined from Worobey *et al.* (1999) or identified by reconstructing bootstrapped phylogenetic trees

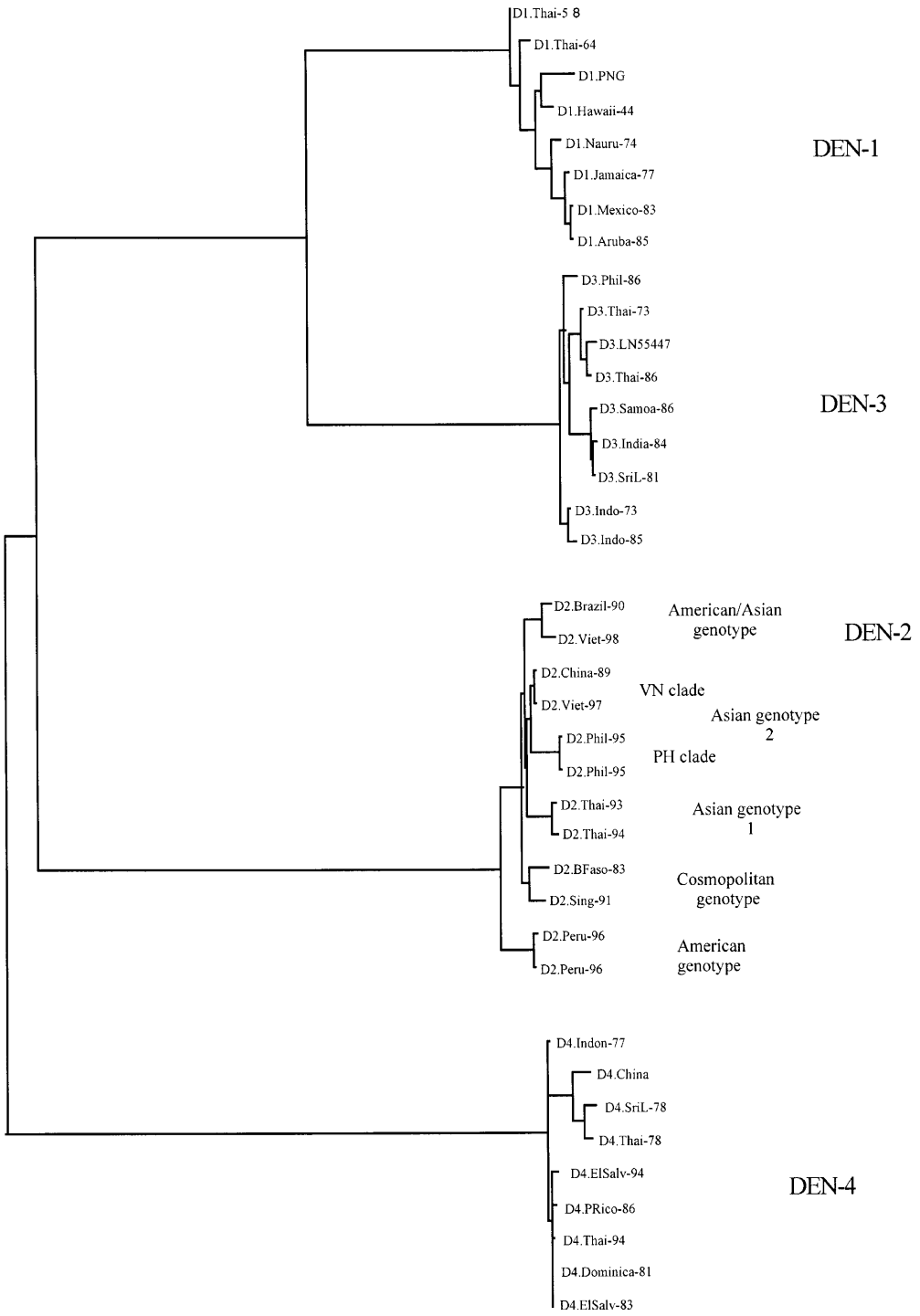

Fig. 1. ML phylogenetic tree depicting the relationships within and among DEN serotypes 1–4 (human strains only). The genotypes of DEN-2 and the two clades within the Asian 2 genotype are labelled. All horizontal branch lengths are drawn to scale and the tree is mid-point-rooted for purposes of clarity only.

from different genomic regions to identify topological shifts. Additionally, in the case of the DEN-2 E gene data set, the large amount of data available meant that sequences with > 99% sequence identity were excluded. It has been demonstrated previously that the removal of closely related sequences in this manner does not bias selection analysis (Yang, 1998).

■ **Phylogenetic analysis.** An ML phylogenetic tree using 38 representative viruses to describe the relationships within and between DEN serotypes is shown in Fig. 1. This tree was constructed using the PAUP* package under the general time-reversible model of nucleotide substitution (Swofford, 2000). Parameter values for each substitution type, optimal base composition, proportion of invariable sites and the

shape parameter (α) of a Γ distribution of rate variation among sites (with eight categories) were estimated from the data and are available from the authors on request. Although each genotype of DEN-2 constitutes a monophyletic group whatever model of sequence evolution is used, the phylogenetic relationships among the genotypes (i.e. the relative positions of the genotypes in the tree topology) are unstable and viruses of Asian 1 genotype are often the first to diverge after the sylvatic strains when trees are reconstructed with DEN-2 isolates alone (Twiddy *et al.*, 2002). For the selection analysis, ML trees were constructed for each data set in question under the HKY85 model of nucleotide substitution, with values for the transition/transversion (T_S/T_V) ratio and α again estimated from the data. All parameter values are available from the authors on request.

■ **Selection analysis.** An ML approach was used to examine selection pressures acting on DEN (Yang *et al.*, 2000; Yang & Bielawski, 2000). In this analysis, ω ratios (d_N/d_S) are determined codon-by-codon using various models of codon substitution that differ in how ω ratios are allowed to vary along the sequence. Six models of codon substitution were used in this study: (1) M0 assumes that all codons are subject to the same selection pressure so that a single ω value is estimated; (2) M1 divides codons into two categories, representing those that are invariant (p_0), with ω_0 fixed at 0, and those (p_1) which are neutral, where ω_1 is set to 1; (3) M2 accounts for positive selection by the addition of a third category of codons (p_2) with ω_2, which can take on any value, including 1, estimated from the data; (4) M3 estimates ω for three classes of codon and provides a more sensitive test for positive selection, as all ω ratios are estimated from the data and all may be 1; (5) M7 uses a discrete β distribution (10 codon classes) to model ω ratios among codons. The β distribution takes on a variety of shapes depending on parameters p and q and, under M7, no class of codons can have an ω ratio > 1; (6) M8 also uses a β distribution but an extra class of codons is incorporated at which ω can be > 1. Models that are nested may be compared statistically using an LRT in which twice the difference in log likelihood between two models is compared to the value obtained under a χ^2 distribution. For the models used here, both M0 and M1 are nested with M2 and M3, and M7 is nested with M8. Positive selection can be inferred when a group of codons with an ω ratio > 1 is identified and the likelihood of the codon substitution model in question is significantly higher ($P < 0{\cdot}05$) than that of a nested model that does not take positive selection into account. Finally, Bayesian methods can be used to calculate the probability that a specific codon falls into the positively selected class. All these methods were implemented using the CODEML program of the PAML package (Yang, 1997).

Results

Because of space limitations, we have only presented results in tabular form where there was evidence of positive selection (Table 2). For this purpose we have, with one exception, shown the results of the analysis using models M3 and M8, as these usually have the highest likelihood and so are the most informative. A full set of results is available at http://evolve.zoo.ox.ac.uk or from the authors on request.

Selection analysis of E gene sequences from all DEN serotypes

No evidence for positive selection was found in the DEN-1 E gene, although there were only a small number of isolates available for analysis. According to the model with the highest likelihood, M3, 92% of sites in this gene are strongly conserved ($\omega = 0{\cdot}036$), with the remainder weakly conserved ($\omega = 0{\cdot}597$). Similarly, no selection was found in the DEN-2 APT data set, where M3 suggested that 72·7% of sites are strongly conserved ($\omega < 0{\cdot}05$), 26% moderately conserved ($\omega < 0{\cdot}50$) and the remaining 1·3% effectively neutral ($\omega = 1{\cdot}051$), nor in the DEN-2 MPO data set, where the best-supported model, M8, suggested that the majority (97%) of sites were not subject to positive selection ($0 < \omega < 1$).

In contrast, for the set of APT DEN-3 strains, both M3 and M8 identified a small class (approximately 1%) of positively selected sites, also agreeing on the strength of selection in this group ($\omega \approx 2{\cdot}1$). Although M3 was unable to unambiguously reject M2, which did not support positive selection, M8 rejected M7, suggesting that the signal is real (Table 2). Under this analysis, only site 380 was assigned to the positively selected class with a probability of $> 90\%$. The results for the MPO data set were very similar, with both models again identifying a small class of positively selected sites ($\sim 1{\cdot}7\%$), with a similar level of selection. As above, M3 was not significantly better than M2, but M8 outperformed M7, again suggesting that the positive selection was genuine. In this case, Bayesian methods identified sites 169 and 380 to be under positive selection.

The analysis of the DEN-4 APT data set provided the strongest evidence for positive selection. In this case, both M3 and M8 indicated that between 1 and 1·5% of codons in the DEN-4 E gene are under positive selection, with ω ratios of 4·368 and 3·999, respectively. Furthermore, both M3 and M8 were significantly better than the corresponding models without positive selection (M2 and M7, respectively). Under M3, sites 108 and 357 were assigned to the positively selected class (probability of $> 90\%$), while under M8, the positively selected sites were 108, 131, 357, 429 and 494. The level of positive selection in the MPO data set was reduced, with M3 and M8 suggesting relatively weak positive selection ($\omega \approx 1{\cdot}8$). Furthermore, neither M3 nor M8 could conclusively reject the models that did not account for positive selection (Table 2).

Selection analysis of the DEN-2 E gene by genotype

Our analysis failed to detect positive selection in the American, American/Asian or Asian 1 genotypes of DEN. In all these cases, most codon positions were strongly conserved. In contrast, in the Asian 2 genotype, both M3 and M8 assigned approximately 3% of sites into a weakly positively selected class ($\omega = 1{\cdot}954$ and $1{\cdot}857$, respectively). However, as neither model was significantly favoured over models M2 or M7, the evidence for positive selection is not conclusive. Because there are two clades within this genotype, which are relatively distinct from one another, comprising viruses predominantly from PH and VN clades, respectively, analyses were carried out on these clades separately. In the case of the

Table 2. DEN data sets where positive selection was detected and their relevant parameter values

The APT data set contains all passage types, whereas the MPO data set contains isolates that were only passaged through mosquito cells. In the case of the DEN-2 E Asian 2 (PH) data set, the M2 model parameters are substituted for the M3 parameters. *P* values in bold indicate comparisons where the null hypothesis can be rejected in favour of the alternative hypothesis (i.e. the model on the left is rejected in favour of the one on the right).

Data set	M3 parameters	M3 LRT	M3 selected sites	M8 parameters	M8 LRT	M8 selected sites
DEN-3 E APT	$p_0 = 0{\cdot}716$ $p_1 = 0{\cdot}272$ $p_2 = 0{\cdot}012$ $\omega_0 = 0{\cdot}001$ $\omega_1 = 0{\cdot}250$ $\boldsymbol{\omega_2 = 2{\cdot}072}$	M1–M3 $\boldsymbol{P < 0{\cdot}001}$ M2–M3 $P = 0{\cdot}456$	380	$p = 0{\cdot}188$ $q = 2{\cdot}281$ $p_0 = 0{\cdot}989$ $p_1 = 0{\cdot}011$ $\boldsymbol{\omega_1 = 2{\cdot}062}$	M7–M8, $\boldsymbol{P < 0{\cdot}001}$	380
DEN-3 E MPO	$p_0 = 0{\cdot}219$ $p_1 = 0{\cdot}764$ $p_2 = 0{\cdot}017$ $\omega_0 = 0{\cdot}053$ $\omega_1 = 0{\cdot}053$ $\boldsymbol{\omega_2 = 1{\cdot}945}$	M1–M3 $\boldsymbol{P < 0{\cdot}001}$ M2–M3 $P = 0{\cdot}379$	169, 380	$p = 1{\cdot}163$ $q = 18{\cdot}827$ $p_0 = 0{\cdot}986$ $p_1 = 0{\cdot}014$ $\boldsymbol{\omega_1 = 2{\cdot}152}$	M7–M8 $\boldsymbol{P = 0{\cdot}028}$	169, 380
DEN-4 E APT	$p_0 = 0{\cdot}942$ $p_1 = 0{\cdot}045$ $p_2 = 0{\cdot}012$ $\omega_0 = 0{\cdot}031$ $\omega_1 = 0{\cdot}840$ $\boldsymbol{\omega_2 = 4{\cdot}368}$	M1–M3 $\boldsymbol{P = 0{\cdot}001}$ M2–M3 $\boldsymbol{P = 0{\cdot}026}$	108, 357	$p = 0{\cdot}120$ $q = 1{\cdot}662$ $p_0 = 0{\cdot}985$ $p_1 = 0{\cdot}015$ $\boldsymbol{\omega_1 = 3{\cdot}999}$	M7–M8 $\boldsymbol{P = 0{\cdot}004}$	108, 131, 357, 429, 494
DEN-4 E MPO	$p_0 = 0{\cdot}509$ $p_1 = 0{\cdot}453$ $p_2 = 0{\cdot}038$ $\omega_0 = 0{\cdot}026$ $\omega_1 = 0{\cdot}031$ $\boldsymbol{\omega_2 = 1{\cdot}799}$	M1–M3 $P = 0{\cdot}067$ M2–M3 $P = 0{\cdot}257$	108, 131, 202, 227, 265, 351, 357, 429	$p = 3{\cdot}984$ $q = 135{\cdot}475$ $p_0 = 0{\cdot}963$ $p_1 = 0{\cdot}037$ $\boldsymbol{\omega_1 = 1{\cdot}810}$	M7–M8 $P = 0{\cdot}061$	108, 131, 202, 227, 265, 351, 357, 429
DEN-2 E Asian 2 (PH)	$p_0 = 0{\cdot}949$ $p_1 = 0{\cdot}000$ $p_2 = 0{\cdot}051$ $\omega_0 = 0{\cdot}000$ $\omega_1 = 1{\cdot}000$ $\boldsymbol{\omega_2 = 2{\cdot}417}$	M0–M2 $\boldsymbol{P < 0{\cdot}001}$ M1–M2 $P = 0{\cdot}069$	52, 85, 90, 98, 100, 105, 112, 113, 122, 131, 144, 170, 330, 334, 342, 378, 392	$p = 0{\cdot}010$ $q = 99{\cdot}000$ $p_0 = 0{\cdot}949$ $p_1 = 0{\cdot}051$ $\boldsymbol{\omega_1 = 2{\cdot}417}$	M7–M8 $P = 0{\cdot}052$	52, 85, 90, 98, 100, 105, 112, 113, 122, 131, 144, 170, 330, 334, 342, 378, 392
DEN-2 E Cosmopolitan	$p_0 = 0{\cdot}784$ $p_1 = 0{\cdot}200$ $p_2 = 0{\cdot}015$ $\omega_0 = 0{\cdot}000$ $\omega_1 = 0{\cdot}213$ $\boldsymbol{\omega_2 = 1{\cdot}805}$	M1–M3 $\boldsymbol{P = 0{\cdot}008}$ M2–M3 $P = 0{\cdot}671$	52, 390	$p = 0{\cdot}119$ $q = 2{\cdot}274$ $p_0 = 0{\cdot}985$ $p_1 = 0{\cdot}015$ $\boldsymbol{\omega_1 = 1{\cdot}800}$	M7–M8 $\boldsymbol{P = 0{\cdot}006}$	52, 390
DEN-2 NS2B	$p_0 = 0{\cdot}098$ $p_1 = 0{\cdot}900$ $p_2 = 0{\cdot}002$ $\omega_0 = 0{\cdot}031$ $\omega_1 = 0{\cdot}031$ $\boldsymbol{\omega_2 = 5{\cdot}137}$	M1–M3 $\boldsymbol{P = 0{\cdot}011}$ M2–M3 $\boldsymbol{P = 0{\cdot}033}$	57, 63	$p = 0{\cdot}029$ $q = 0{\cdot}512$ $p_0 = 1{\cdot}000$ $p_1 = 0{\cdot}000$ $\omega_1 = \mathrm{NA}$	M7–M8 $P = 1{\cdot}000$	None
DEN-2 NS5	$p_0 = 0{\cdot}085$ $p_1 = 0{\cdot}914$ $p_2 = 0{\cdot}001$ $\omega_0 = 0{\cdot}439$ $\omega_1 = 0{\cdot}005$ $\boldsymbol{\omega_2 = 4{\cdot}187}$	M1–M3 $\boldsymbol{P < 0{\cdot}001}$ M2–M3 $\boldsymbol{P = 0{\cdot}012}$	135, 637	$p = 0{\cdot}012$ $q = 0{\cdot}237$ $p_0 = 0{\cdot}999$ $p_1 = 0{\cdot}001$ $\boldsymbol{\omega_1 = 4{\cdot}101}$	M7–M8 $\boldsymbol{P < 0{\cdot}001}$	135, 637

PH clade, the M2 and M8 models had the highest likelihoods and indicated that 5% of sites may be subject to positive selection ($\omega = 2{\cdot}419$ for M2 and 2·417 for M8), while Bayesian methods assigned the same 17 codons to the positively selected class with $> 99\%$ probability for both models (Table 2). Although M2 and M8 did not reject M1 and M7 unequivocally ($P = 0{\cdot}069$ for M2 versus M1 and 0·052 for M8 versus M7), we judged that these P values were sufficiently borderline for the sites identified as subject to positive selection to merit further investigation. In contrast, the best-supported model in the analysis of the VN clade was M0, which indicated that all sites are moderately conserved ($\omega = 0{\cdot}377$).

Positive selection was also detected in the geographically widespread Cosmopolitan genotype. The model with the highest likelihood in this analysis was M3, but both M3 and M8 identified a small class of positively selected sites (1·5%, $\omega = 1{\cdot}805$ and 1·800, respectively). Although M3 was not able to reject competing neutral models of evolution at the 95% level, the P value for M8 versus M7 comparison (0·006) was highly significant, suggesting that positive selection was indeed operating within this genotype. The sites assigned to the positively selected class with $> 90\%$ probability under both M3 and M8 were codon positions 52 and 390.

Selection analysis of other genes of DEN-2

No evidence of adaptive evolution was found in the two other structural genes of DEN-2, encoding the capsid (C) and the premembrane/membrane (prM/M) proteins, with strong purifying selection the predominant evolutionary pressure ($\omega < 0{\cdot}001$ for 90% of amino acid sites in C and 88–98% of sites in prM/M).

Of the seven non-structural genes, five showed no evidence for positive selection. In the NS1 data set, the best-supported model, M3, suggested that 79% of sites are completely conserved, with the remainder moderately conserved. The results were similar for NS2A, with both M2 and M3 suggesting that 96–97% of sites are strongly conserved ($\omega = 0{\cdot}032$ and 0·031 for M2 and M3, respectively), and the other 3–4% are neutral ($\omega = 1$ and 0·982 for M2 and M3, respectively). The majority of sites in NS3 are also strongly conserved, with just over 2% being effectively neutral ($\omega = 0{\cdot}856$ under the highest likelihood model M3). Models M2 and M3 were equally well supported in the analysis of the NS4A gene, both suggesting that 85% of sites are completely conserved and 15% moderately conserved. Finally, the best-performing models for the NS4B data set, M3 and M8, agreed that 97·3% of sites are strongly or completely conserved, while the remaining 2·7% are close to neutral ($\omega > 0{\cdot}9$).

In contrast, positive selection was detected in both NS2B and NS5. For the NS2B data set, M3 was significantly better than M2 and indicated that 0·2% of sites are subject to relatively strong positive selection ($\omega = 5{\cdot}137$). However, there was no evidence for positive selection under M8. Under M3, codon positions 57 and 63 were assigned to the positively selected class. In the NS5 data set, both M3 and M8 identified a small (1·0%) class of positively selected sites ($\omega = 4{\cdot}187$ and 4·101 for M3 and M8, respectively), which corresponds to codon positions 135 and 637. In addition, both M3 and M8 unanimously rejected models that did not account for positive selection, indicating that the signal for adaptive evolution in this data set is relatively strong.

Discussion

Our analysis provides evidence for the positive selection of amino acid positions in DEN, although on a highly localized basis. We now discuss the possible selection pressures at each of these sites in turn (summarized in Table 3).

Selection pressures within the E gene of DEN

No evidence was found for positive selection in either of the complete data sets from DEN-1 and DEN-2, although some selection was found in individual genotypes of DEN-2 (see below). In contrast, adaptive evolution was detected in DEN-3 and, most strongly, DEN-4. These results imply that different DEN serotypes are subject to different selection pressures for reasons that are unclear. However, the positive selection found was relatively weak and it is possible that a denser sample of sequences may reveal more widespread evidence of adaptive evolution.

DEN-3. Two positively selected sites were identified in the MPO data set of DEN-3. The first – aa 169 – is located in both murine T- and B-cell epitopes (Roehrig *et al.*, 1994; Leclerc *et al.*, 1993). This suggests that it may be subject to immune selection, although care must be taken when extrapolating results from animal studies to the human experience. The second selected site – aa E-380 – is located on the distal face of domain III (receptor-binding domain) of the glycoprotein, the structure of which resembles an immunoglobulin and within which mutations may affect cell attachment (Rey *et al.*, 1995). Residue E-380 is also located within a 4 bp motif that is absent in the tick-borne flaviviruses, although relatively highly variable among the DEN complex (Gritsun *et al.*, 1995; McMinn, 1997). These observations are compatible with a role in cell tropism.

DEN-4. Five amino acid sites in the E gene of DEN-4 were found to be subject to positive selection. The selection at two (E-357 and E-429) is likely to involve T- or B-cell epitopes. It has been suggested that the region encompassing aa 333–368 is an immunodominant region containing multiple B- and T-cell epitopes and peptides from this region have been shown to bind to homologous and heterologous sera from DEN patients (Innis *et al.*, 1989), elicit virus-binding antibody, stimulate T-cell proliferation in mice and react strongly with DEN cross-reactive monoclonal antibodies (Aaskov *et al.*, 1989; Roehrig *et al.*, 1994; Falconar, 1999). Similarly, the region immediately surrounding residue E-429 has been predicted on the basis of

its secondary structure to be a flavivirus cross-reactive Th-cell epitope (Kutubuddin *et al.*, 1991).

Two of the other selected sites in DEN-4, residues E-108 and E-131, fall in regions where mutations would be expected to alter membrane fusion properties. The highly conserved flavivirus 'fusion peptide', which interacts with the host endosomal membrane, leading to virus-mediated membrane fusion and allowing the newly infecting virus to initiate the cellular replication cycle, is thought to comprise aa 98–111 (Mandl *et al.*, 1989; Roehrig *et al.*, 1990). As residue E-108 is located within the fusion peptide, it is likely to directly affect the process of membrane fusion. According to a structural model of the flavivirus E glycoprotein (Rey *et al.*, 1995), residue E-131 is also located in a region of the protein where mutations are expected to affect virus-mediated membrane fusion. However, in this case, the effect is indirect, via the low-pH conformational change, which exposes the fusion peptide on the outer surface of the virion, allowing interaction with the host membrane (Roehrig *et al.*, 1994).

The final selected site in DEN-4, E-494, is at the extreme C-terminal region of the E glycoprotein, immediately following the transmembrane NS1 signal sequence (Chang, 1996). A peptide encompassing aa 470–493 was used by Roehrig *et al.* (1998) as a control peptide in antibody-binding assays, suggesting that this region is probably not immunogenic. It has been proposed (Wang *et al.*, 1999) that the C-terminal portion of the E protein may be the location of prM-binding sites, stabilizing the E–prM network within the virus particle, although the precise locations of these sites have not been defined. However, it is also known that the amino acid composition of the final three residues at the C terminus of the protein is constrained by the '−1, −3 rule' for signal peptidase cleavage, which requires Ala, Ser, Gly, Cys, Thr or Gln in position −1 (E-495 in this case), exclusion of aromatic, charged or polar residues from position −3 (E-493) and the absence of Pro in positions −3 to +1 (Biedrzycka *et al.*, 1987). The substitution at the selected site conforms to this rule (H → Q at E-494). It is difficult to assess what selective advantage this substitution could give to the mutant strains, although it is possible that it increases the efficiency of virus processing.

Selection within genotypes of DEN-2

It has been proposed that the American genotype may represent a low virulence genotype of DEN-2 and that the amino acid difference between this and other DEN-2 genotypes at E-390 (N → D) may be critically involved in determining virulence (Leitmeyer *et al.*, 1999). Our analysis found no evidence for selection at this or any other sites in the American genotype but as all residues at E-390 are identical within the American genotype, an analysis based on variable codon positions would be unable to detect selection at this site. Furthermore, E-390 was found to be under positive selection in another genotype of DEN-2 (see below).

Similarly, we found no evidence for positive selection in the American/Asian or Asian 1 genotypes. The results for the Asian 2 genotype, however, suggested that there might be selection within the PH subclade, with 17 amino acid sites identified as selected, although with borderline significance levels. Of these, 12 – positions 52, 85, 90, 122, 131, 144, 170, 330, 334, 342, 378 and 392 – are candidates for immune selection, with synthetic peptides encompassing each of these sites showing T- and/or B-cell reactivity (Aaskov *et al.*, 1989; Innis *et al.*, 1989; Roehrig *et al.*, 1994; Megret *et al.*, 1992; Leclerc *et al.*, 1993). Moreover, selection at three sites – E-342, E-378 and E-392 – is likely to involve cell tropism, as all are located on the distal face of domain III (Rey *et al.*, 1995) within a region that shows a different pattern of amino acid variability between DEN and the tick-borne flaviviruses (Gritsun *et al.*, 1995) and which also contains multiple residues that differ between flaviviruses according to their vectors (McMinn, 1997). Finally, seven sites – 52, 98, 100, 105, 112, 113 and 131 – may be under positive selection due to their ability to affect virus-mediated membrane fusion and therefore virus replication. This may either be a direct effect, for those sites which fall within or adjacent to the fusion peptide, or indirect, via the low-pH conformational change. The identification of selection for an amino acid other than Cys at site 105 is particularly intriguing. Cys at position 105 is known to be part of a disulphide bridge that is thought to stabilize the 'cd loop' structure that the fusion peptide adopts (Rey *et al.*, 1995). As the same mutation (C → F) appears three times within the group of 12 strains, it is unlikely to be sequencing error. What effect this substitution might have on protein structure in this crucial region is unknown but strains carrying this mutation have been successfully passaged in mosquito cells (R. Matias, personal communication). It is also uncertain why there should be extensive positive selection within this clade, which is found most often in the Philippines, although this clearly merits further study.

In contrast to the PH clade, the Cosmopolitan genotype, which was also found to be subject to positive selection, has a near global geographical distribution (Twiddy *et al.*, 2002). Two amino acid sites may be selected in this genotype. The first of these, E-52, was also found to be under selection in the PH group of DEN-2 and, as discussed above, may affect virus-mediated membrane fusion by affecting the low-pH conformational change. In addition, there is evidence for a B-cell epitope in this region (Aaskov *et al.*, 1989, Roehrig *et al.*, 1994). The second, and more interesting, selected site in the Cosmopolitan genotype is at codon position E-390, with the amino acid replacement N → S present in 21 of 28 strains. This residue is located within domain III, which, as described previously, may be involved in cell receptor binding. Furthermore, substitutions at residue 390 have been shown to alter DEN-2 neurovirulence in mice using a strain belonging to the American genotype, with a D → H mutation increasing virulence and a D → N replacement leading to attenuation (Sanchez & Ruiz, 1996). It

Table 3. Correlation of positively selected sites in DEN with known biological features

Serotype/ genotype	Gene	Selected site	Amino acid substitution	Possible selection pressure	Comment
DEN-3	E	169	V → T/A	Immune selection	Peptide 142–172 T- and B-cell reactive Peptide 163–182 dominant T-cell epitope
		380	I → T/S	Cell tropism	Located on distal face of receptor-binding domain Within 4 bp motif exclusive to mosquito-borne flaviviruses
DEN-4	E	108	F → L	Alteration of membrane fusion properties	Part of strongly conserved fusion peptide Substitutions in DEN-1 resulted in different membrane fusion properties
		131	Q → R	Alteration of fusion properties/ Immune selection	Peptide 49–60 + 121–40 strongly B-cell reactive at low pH only Mutations at this site may affect the low-pH conformational change Peptide 127–134 recognized by sera from convalescent DEN-2 patients
		357	I → V	Immune selection	Immunodominant region 333–368 – multiple T- and B-cell epitopes
		429	F → S/L		Predicted Th-cell epitope
		494	H → Q	Unknown (efficiency of virus processing?)	Conforms to '−1, −3' rule for signalase cleavage (−2 position)
DEN-2 Asian 2 (PH)	E	52	Q → E	Immune selection/alteration of fusion properties	Peptide 35–55 T- and B-cell reactive Peptide 49–60 + 121–40 strongly B-cell reactive at low pH Mutations at this site may affect the low-pH conformational change
		85	E → K	Immune selection	Peptide 79–99 T- and B-cell reactive
		90	F → S	Immune selection	Peptide 79–99 T- and B-cell reactive
		98	D → N	Alteration of fusion properties	Part of fusion peptide
		100	G → R	Alteration of fusion properties	Part of fusion peptide
		105	C → F	Alteration of fusion properties	Part of fusion peptide Involved in Cys = Cys disulphide bridge
		112	G → S	Alteration of fusion properties	Adjacent to fusion peptide.
		113	I → V	Alteration of fusion properties	Adjacent to fusion peptide.
		122	K → Q	Immune selection	Part of cryptic T-cell epitope.
		131	Q → P	Alteration of fusion properties/ immune selection	Peptide 49–60 + 121–40 strongly B-cell reactive at low pH only Mutations at this site may affect the low-pH conformational change Peptide 127–134 recognized by sera from convalescent DEN-2 patients
		144	H → P	Immune selection	Peptide 135–157 contained Th-cell epitope
		170	I → T		Peptide 142–172 was both T- and B-cell reactive Peptide 165–201 stimulated Th-cell response, without antibody response
		330	G → D	Immune selection	Most synthetic peptides from region 301–388 show T- and B-cell reactivity
		334	K → Q		
		342	L → W	Cell tropism/immune selection	
		378	I → V		Region of domain III encompassing these sites (distal face) is highly variable in DEN but conserved in the tick-borne flaviviruses Multiple sites in this region differ between flaviviruses according to their vectors Peptide 386–392 bound homologous and heterologous human DEN sera
		392	F → Y		

Table 3 (*cont.*)

Serotype/ genotype	Gene	Selected site	Amino acid substitution	Possible selection pressure	Comment
DEN-2 Cosmopolitan	E	52	Q → H	Immune selection/alteration of fusion properties	Peptide 35–55 T- and B-cell reactive Peptide 49–60 + 121–40 strongly B-cell reactive at low pH only Mutations at this site may affect the low-pH conformational change
		390	N → S	Cell tropism	In domain III (receptor-binding) Substitutions at this site have been implicated in determining virulence
DEN-2	NS2B	57	T → A	Unknown	Within 40 aa segment of protein essential for NS2B/NS3 protease activity
		63	D → R/N	Unknown	
DEN-2	NS5	135	I → T/V	Unknown	No information available
		637	A → V/S/I	Unknown	May be involved in NS5 polymerase activity

is also significant that E-390 has been identified as a residue that may determine some characteristics of the American genotype, with all these viruses showing an N → D amino acid replacement at this site (Leitmeyer *et al.*, 1999). In summary, these observations suggest that the character of the amino acid at E-390 may indeed be important in determining key aspects of virus phenotype, although this clearly requires further investigation.

Positive selection in other structural and non-structural proteins in DEN-2

There was no evidence of positive selection in the structural proteins of DEN-2, with the exception of the E glycoprotein. Likewise, no evidence for positive selection was found in the non-structural proteins NS1, NS2A, and NS3, despite plentiful evidence for the existence of numerous T- and B-cell epitopes in these genes (Henchal *et al.*, 1987; Kurane *et al.*, 1991; Falconar *et al.*, 1994; Garcia *et al.*, 1997; Spaulding *et al.*, 1999), nor in NS4A or NS4B. However, there was evidence for selection in the small and relatively little-studied protein NS2B. The only known function of this protein is to act as cofactor for the viral protease NS3, which generates the N-termini of NS2B, NS3, NS4A and NS5. It is probable that NS2B interacts with NS3 to form a complex that maintains the polyprotein precursor in a conformation that NS3 is able to cleave (Falgout *et al.*, 1991; Roehrig, 1996) and that this interaction is mediated by a 40 aa segment of NS2B that has been shown to be essential for NS2B/NS3 protease activity (Falgout *et al.*, 1993). Both positively selected amino acid replacements identified in the DEN-2 NS2B data set (NS2B-57, T → A, and NS2B-63, D → R/N) fall within this region and may therefore benefit virus strains by increasing the efficiency of polyprotein processing.

There was also evidence for positive selection in the genes encoding the NS5 protein. Whilst it is well established that NS5 is required for virus replication (Bartholomeusz & Wright, 1993), little else is known about the structure and function of this, the largest of the DEN proteins. The N-terminal region of the flavivirus NS5 protein contains a sequence motif that is conserved in *S*-adenosylmethionine-utilizing methyltransferases (Forwood *et al.*, 1999) and may therefore be involved in virus capping. The C-terminal region (from residue 455 onwards) contains motifs that bear similarity to those found in known RNA-dependent RNA polymerases (RdRp) (Raviprakash *et al.*, 1998) and bacterially expressed NS5 has been shown to possess RdRp activity. However, neither of the two amino acid sites that we determined to be under positive selection – NS5-135 and NS5-637 – fell into functionally defined regions, although NS5-637 is near the GDD motif found in most RdRps and may be involved in NS5 polymerase activity. Consequently, the phenotypic importance of these mutations is unclear.

Overall, our analysis reveals that there are multiple selected amino acid positions in the genes of DEN, in contrast to the observations of previous studies (Yang *et al.*, 2000; Zanotto *et al.*, 1996). This might be expected in the case of the E glycoprotein given its role as the major surface antigen of the virus. However, in similar analyses of E gene sequences from other flaviviruses, including Japanese encephalitis virus, St Louis encephalitis virus, West Nile virus and yellow fever virus, no positive selection has been detected (C. Woelk, personal communication; Yang *et al.*, 2000). Why DEN should differ from other flaviviruses in this respect is not known. However, it is equally clear that the selection pressures acting

on DEN are relatively weak when compared to those in viruses seemingly subject to strong host immune pressure, viruses such as influenza A virus, hepatitis C virus and human immunodeficiency virus (Bush *et al.*, 1999; Farci *et al.*, 2000; Zanotto *et al.*, 1999). It is therefore possible that the more complex nature of the virus life-cycle in DEN, involving vertebrate and invertebrate hosts, both of which would produce a substantial bottleneck at transmission and perhaps impose stronger selective constraints than on directly transmitted viruses, mediates the strength of selection acting on this and other vector-borne viruses.

While most of the amino acid substitutions under selection could be mapped to epitopes recognized by components of the vertebrate cellular and humoral immune response, several others appear to affect the tertiary structure of the E glycoprotein and so may also have implications for cell tropism, via modification of receptor-binding sites, and virus-mediated membrane fusion, through modification of the fusion peptide or the low-pH conformational change. It is also possible that some of the observed changes could be due to adaptation to mosquito vectors as well as to vertebrate hosts and there is evidence that strains of DEN-2 may differ in their ability to infect *Aedes aegypti* mosquitoes (Armstrong & Rico-Hesse, 2002). We suggest that all the putatively selected sites identified here should now be the subjects of further experimental study.

This work was supported by research grants from The Wellcome Trust, The Royal Society and the BBSRC.

References

Aaskov, J. G., Geysen, H. M. & Mason, T. J. (1989). Serologically defined linear epitopes in the envelope protein of dengue 2 (Jamaica strain 1409). *Archives of Virology* **105**, 209–221.

Armstrong, P. M. & Rico-Hesse, R. (2002). Differential susceptibility of *Aedes aegypti* to infection by the American and Southeast Asian genotypes of dengue type 2 virus. *Vector Borne and Zoonotic Diseases* **1**, 159–168.

Bartholomeusz, A. I. & Wright, P. J. (1993). Synthesis of dengue virus RNA *in vitro*: initiation and the involvement of proteins NS3 and NS5. *Archives of Virology* **128**, 111–121.

Biedrzycka, A., Cauchi, M. R., Bartholomeusz, A., Gorman, J. J. & Wright, P. J. (1987). Characterization of protease cleavage sites involved in the formation of the envelope glycoprotein and three non-structural proteins of dengue virus type 2, New Guinea C strain. *Journal of General Virology* **68**, 1317–1326.

Bush, R. M., Fitch, W. M., Bender, C. A. & Cox, N. J. (1999). Positive selection on the H3 hemagglutinin gene of human influenza virus A. *Molecular Biology and Evolution* **16**, 1457–1465.

Chang, G.-J. (1996). Molecular biology of dengue viruses. In *Dengue and Dengue Hemorrhagic Fever*, pp. 175–198. Edited by D. J. Gubler & G. Kuno. New York: CAB International.

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. *Journal of Virology* **74**, 7814–7823.

Falconar, A. K. I. (1999). Identification of an epitope on the dengue virus membrane (M) protein defined by cross-protective monoclonal antibodies: design of an improved epitope sequence based on common determinants present in both envelope (E and M) proteins. *Archives of Virology* **144**, 2313–2330.

Falconar, A. K. I., Young, P. R. & Miles, M. A. (1994). Precise location of sequential dengue virus subcomplex and complex B cell epitopes on the nonstructural-1 glycoprotein. *Archives of Virology* **137**, 315–326.

Falgout, B., Pethel, M., Zhang, Y.-M. & Lai, C.-J. (1991). Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. *Journal of Virology* **65**, 2467–2475.

Falgout, B., Miller, R. H. & Lai, C.-J. (1993). Deletion analysis of dengue virus type 4 nonstructural protein NS2B: identification of a domain required for NS2B–NS3 protease activity. *Journal of Virology* **67**, 2034–2042.

Farci, P., Shimoda, A., Coiana, A., Diaz, G., Peddis, G., Melpolder, J. C., Strazzera, A., Chien, D. Y., Munoz, S. J., Balestrieri, A., Purcell, R. H. & Alter, H. J. (2000). The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* **288**, 339–344.

Forwood, J. K., Brooks, A., Briggs, L. J., Xiao, C. Y., Jans, D. A. & Vasudevan, S. G. (1999). The 37-amino-acid interdomain of dengue virus NS5 protein contains a functional NLS and inhibitory CK2 site. *Biochemical and Biophysical Research Communications* **257**, 731–737.

Garcia, G., Vaughn, D. W. & del Angel, R. M. (1997). Recognition of synthetic oligopeptides from nonstructural proteins NS1 and NS3 of dengue-4 virus by sera from dengue virus-infected children. *American Journal of Tropical Medicine and Hygiene* **56**, 466–470.

Gritsun, T. S., Holmes, E. C. & Gould, E. A. (1995). Analysis of flavivirus envelope proteins reveals variable domains that reflect their antigenicity and may determine their pathogenesis. *Virus Research* **35**, 307–321.

Henchal, E. A., McCown, J. M., Burke, D. S., Seguin, M. C. & Brandt, W. E. (1985). Epitopic analysis of antigenic determinants on the surface of dengue-2 virions using monoclonal antibodies. *American Journal of Tropical Medicine and Hygiene* **34**, 162–169.

Henchal, E. A., Henchal, L. S. & Thaisomboonsuk, B. K. (1987). Topological mapping of unique epitopes on the dengue-2 virus NS1 protein using monoclonal antibodies. *Journal of General Virology* **68**, 845–851.

Innis, B. L., Thirawuth, V. & Hemachudha, C. (1989). Identification of continuous epitopes of the envelope glycoprotein of dengue type 2 virus. *American Journal of Tropical Medicine and Hygiene* **40**, 676–687.

Kurane, I., Brinton, M. A., Samson, A. L. & Ennis, F. A. (1991). Dengue virus-specific, human $CD4^+CD8^-$ cytotoxic T-cell clones: multiple patterns of virus cross-reactivity recognized by NS3-specific T-cell clones. *Journal of Virology* **65**, 1823–1828.

Kutubuddin, M., Kolaskar, A. S., Galande, S., Gore, M. M., Ghosh, S. N. & Banerjee, K. (1991). Recognition of helper T cell epitopes in envelope (E) glycoprotein of Japanese encephalitis, West Nile and dengue viruses. *Molecular Immunology* **28**, 149–154.

Leclerc, C., Dériaud, E., Megret, F., Briand, J.-P., van Regenmortel, M. H. V. & Deubel, V. (1993). Identification of helper T cell epitopes of dengue virus E-protein. *Molecular Immunology* **30**, 613–625.

Leitmeyer, K. C., Vaughn, D. W., Watts, D. M., Salas, R., Villalobos, I., de Chacon, I. V., Ramos, C. & Rico-Hesse, R. (1999). Dengue virus structural differences that correlate with pathogenesis. *Journal of Virology* **73**, 4738–4747.

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.

McMinn, P. C. (1997). The molecular basis of virulence of the encephalitogenic flaviviruses. *Journal of General Virology* **78**, 2711–2722.

Mandl, C. W., Holzmann, H., Guirakhoo, F., Tuma, W., Heinz, F. X. & Kunz, C. (1988). Antigenic structure of the flavivirus envelope protein. *Gemeinsame Herbsttagung* **369**, 872.

Mandl, C. W., Guirakhoo, F., Holzmann, H., Heinz, F. X. & Kunz, C. (1989). Antigenic structure of the flavivirus envelope protein E at the molecular level, using tick-borne encephalitis virus as a model. *Journal of Virology* **63**, 564–571.

Megret, F., Hugnot, J. P., Falconar, A., Gentry, M. K., Morens, D. M., Murray, J. M., Schlesinger, J. J., Wright, P. J., Young, P., van Regenmortel, M. H. V. & Deubel, V. (1992). Use of recombinant fusion proteins and monoclonal antibodies to define linear and discontinuous antigenic sites on the dengue virus envelope glycoprotein. *Virology* **187**, 480–491.

Raviprakash, K., Sinha, M., Hayes, C. G. & Porter, K. R. (1998). Conversion of dengue virus replicative form RNA (RF) to replicative intermediate (RI) by nonstructural proteins NS-5 and NS-3. *American Journal of Tropical Medicine and Hygiene* **58**, 90–95.

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., 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. *American Journal of Tropical Medicine and Hygiene* **58**, 96–101.

Roehrig, J. T. (1996). Immunochemistry of dengue viruses. In *Dengue and Dengue Hemorrhagic Fever*, pp. 199–219. Edited by D. J. Gubler & G. Kuno. New York: CAB International.

Roehrig, J. T., Johnson, A. J., Hunt, A. R., Bolin, R. A. & Chu, M. C. (1990). Antibodies to dengue 2 virus E-glycoprotein synthetic peptides identify antigenic conformation. *Virology* **177**, 668–675.

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.

Roehrig, J. T., Bolin, R. A. & Kelly, R. G. (1998). Monoclonal antibody mapping of the envelope glycoprotein of the dengue 2 virus, Jamaica. *Virology* **246**, 317–328.

Sanchez, I. J. & Ruiz, B. H. (1996). A single nucleotide change in the E protein gene of dengue virus 2 Mexican strain affects neurovirulence in mice. *Journal of General Virology* **77**, 2541–2545.

Spaulding, A. C., Kurane, I., Ennis, F. A. & Rothman, A. L. (1999). Analysis of murine $CD8^+$ T-cell clones specific for the dengue virus NS3 protein: flavivirus cross-reactivity and influence of infecting serotype. *Journal of Virology* **73**, 398–403.

Swofford, D. L. (2000). PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods), version 4. Sinauer Associates. Sunderland, MA, USA.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Twiddy, S. S., Farrar, J. F., Chau, N. V., Wills, B., Gould, E. A., Gritsun, T., Lloyd, G. & Holmes, E. C. (2002). Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus. *Virology* (in press).

Wang, S., He, R. & Anderson, R. (1999). PrM- and cell-binding domains of the dengue virus E protein. *Journal of Virology* **73**, 2547–2551.

WHO (2000). Strengthening implementation of the global strategy for dengue fever/dengue haemorrhagic fever prevention and control: report of the informal consultation (WHO Headquarters, Geneva, 18–20 October 1999; http://www.who.int/emc-documents/dengue/whocdsdenic20001c.html).

Woelk, C. H., Jin, L., Holmes, E. C. & Brown, D. W. G. (2001). Immune and artificial selection in the haemagglutinin (H) glycoprotein of measles virus. *Journal of General Virology* **82**, 2463–2474.

Worobey, M., Rambaut, A. & Holmes, E. C. (1999). Widespread intraserotype recombination in natural populations of dengue virus. *Proceedings of the National Academy of Sciences, USA* **96**, 7352–7357.

Yang, Z. H. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* **13**, 555–556.

Yang, Z. H. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution* **15**, 568–573.

Yang, Z. & Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution* **15**, 496–503.

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

Zanotto, P. M., Gould, E. A., Gao, G. F., Harvey, P. H. & Holmes, E. C. (1996). Population dynamics of flaviviruses revealed by molecular phylogenies. *Proceedings of the National Academy of Sciences, USA* **93**, 548–553.

Zanotto, P. M., Kallas, E. G., de Souza, R. F. & Holmes, E. C. (1999). Genealogical evidence for positive selection in the *nef* gene of HIV-1. *Genetics* **153**, 1077–1089.

Received 14 December 2001; Accepted 10 March 2002