

# Molecular evolution of dengue 2 virus in Puerto Rico: positive selection in the viral envelope accompanies clade reintroduction

Shannon N. Bennett,<sup>1†</sup> Edward C. Holmes,<sup>2</sup> Maritza Chirivella,<sup>3‡</sup> Dania M. Rodriguez,<sup>1</sup> Manuela Beltran,<sup>3</sup> Vance Vorndam,<sup>3</sup> Duane J. Gubler<sup>4</sup> and W. Owen McMillan<sup>1</sup>

## Correspondence

Shannon N. Bennett  
sbennett@hawaii.edu

<sup>1</sup>Department of Biology, University of Puerto Rico – Rio Piedras, San Juan, PR, USA

<sup>2</sup>Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, PA 16802, USA

<sup>3</sup>Centers for Disease Control and Prevention, Dengue Branch, San Juan, PR, USA

<sup>4</sup>Asia-Pacific Institute of Tropical Medicine and Infectious Diseases, Honolulu, HI, USA

*Dengue virus* is a circumtropical, mosquito-borne flavivirus that infects 50–100 million people each year and is expanding in both range and prevalence. Of the four co-circulating viral serotypes (DENV-1 to DENV-4) that cause mild to severe febrile disease, DENV-2 has been implicated in the onset of dengue haemorrhagic fever (DHF) in the Americas in the early 1980s. To identify patterns of genetic change since DENV-2's reintroduction into the region, molecular evolution in DENV-2 from Puerto Rico (PR) and surrounding countries was examined over a 20 year period of fluctuating disease incidence. Structural genes (over 20% of the viral genome), which affect viral packaging, host-cell entry and immune response, were sequenced for 91 DENV-2 isolates derived from both low- and high-prevalence years. Phylogenetic analyses indicated that DENV-2 outbreaks in PR have been caused by viruses assigned to subtype IIIb, originally from Asia. Variation amongst DENV-2 viruses in PR has since largely arisen *in situ*, except for a lineage-replacement event in 1994 that appears to have non-PR New World origins. Although most structural genes have remained relatively conserved since the 1980s, strong evidence was found for positive selection acting on a number of amino acid sites in the envelope gene, which have also been important in defining phylogenetic structure. Some of these changes are exhibited by the multiple lineages present in 1994, during the largest Puerto Rican outbreak of dengue, suggesting that they may have altered disease dynamics, although their functional significance will require further investigation.

Received 1 July 2005

Accepted 4 December 2005

## INTRODUCTION

*Dengue virus* (DENV) is a mosquito-borne RNA virus (genus *Flavivirus*, family *Flaviviridae*) with an expanding circum-tropical distribution, whose health impact on human populations has become increasingly severe in recent decades (WHO, 1999; Gubler, 2002). DENV causes disease in humans, with symptoms ranging from mild fever to potentially fatal

dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS). Today's increased health burden reflects a greater incidence of severe cases, in addition to more frequent epidemics and geographical expansion into new areas (Gubler, 1998a, 2002). Many of these epidemiological changes can be attributed to the increase and spread of the vector *Aedes aegypti*, an urban mosquito, along with growth, urbanization and travel amongst human populations (Gubler, 1998a). However, populations of DENV are also genetically diverse and there is some evidence for adaptive evolution (Lewis *et al.*, 1993; Rico-Hesse *et al.*, 1997, 1998; Wang *et al.*, 2000; Twiddy *et al.*, 2002b; Bennett *et al.*, 2003) that, in one case, has been correlated with disease incidence (e.g. numbers of cases and isolates; Bennett *et al.*, 2003). Thus, an examination of epidemiological change in DENV incidence must simultaneously consider changes in virus genomes.

†Present address: Asia-Pacific Institute of Tropical Medicine and Infectious Diseases, University of Hawaii at Manoa, 651 Ilalo St, BSB 320, Honolulu, HI 96813, USA.

‡Present address: Amgen Manufacturing Ltd, Juncos, PR, USA.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are DQ364476–DQ364566.

A supplementary table with details of the isolates used in this study is available in JGV Online.

DENV consists of four closely related serotypes (DENV-1 to DENV-4) that were first described based on unique host-antibody responses, but are supported by considerable genetic divergence since their Asian/African origins in sylvatic primates (Gubler, 1997; Wang *et al.*, 2000). These can be further subdivided into subtypes (also known as 'subtypes' or 'genotypes') based on levels of genetic variation (Trent *et al.*, 1989; Rico-Hesse, 1990; Lewis *et al.*, 1993) and, although all four serotypes can produce severe disease, certain subtypes such as Asian DENV-2 have been associated more often with DHF/DSS (Rico-Hesse *et al.*, 1997; Gubler, 1998a; Leitmeyer *et al.*, 1999). However, it has been difficult to associate patterns of viral genetic variation with phenotypic changes in disease for several reasons: (i) DENV does not cause similar disease in animal or *in vitro* models; (ii) phenotypic changes in transmission or virulence are confounded with changing epidemiology due to host factors [e.g. increasing numbers of severe cases have been partly attributed to enhancement of infection in hosts with immunity to other serotypes (Halstead, 1988; Kliks *et al.*, 1989; Thein *et al.*, 1997)]; and (iii) few studies are restricted to a single host population with sufficient temporal sampling and supporting epidemiological data to associate molecular evolution with phenotypic change.

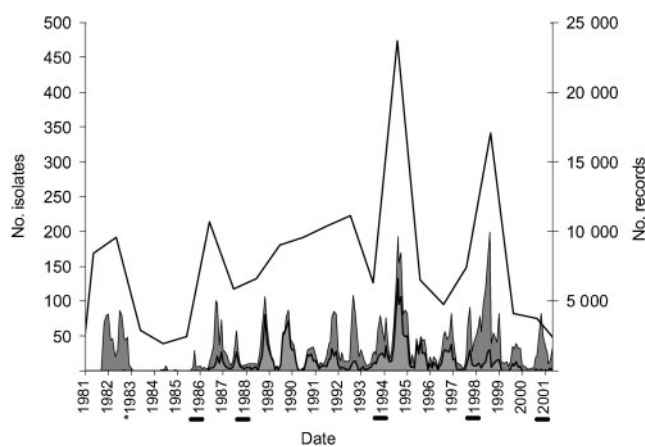
Dengue expansion in the Americas has been particularly evident in Puerto Rico (PR), which provides an opportunity for detailed phylogenetic study of dengue evolution over the last two decades. A densely populated island in the Caribbean, PR first experienced a major dengue outbreak in 1915 and DENV-2 was first identified in 1969 (Dietz *et al.*, 1996). However, PR has only experienced continuous dengue transmission of multiple serotypes since 1985 (Gubler, 1993), along with steadily larger and more frequent epidemics marked by increased numbers of DHF/DSS cases (Dietz *et al.*, 1996; Gubler, 1998b; WHO, 1999). The first major DHF/DSS epidemic in PR occurred in 1986, the year following hyperendemic transmission, where three serotypes were present (DENV-1, -2 and -4), each associated with DHF/DSS (Dietz *et al.*, 1996). Hyperendemic transmission in the Americas began somewhat earlier (Gubler, 1993), but the first DHF/DSS epidemic in the region was in Cuba in 1981 (Kouri *et al.*, 1989) and was associated specifically with a DENV-2 subtype new to the Americas (Rico-Hesse, 1990; Lewis *et al.*, 1993; Guzman *et al.*, 1995; Rico-Hesse *et al.*, 1997). DENV-2 in the Americas prior to 1981 had only been associated with classic dengue fever, despite co-circulation with other serotypes (Gubler, 1997; Watts *et al.*, 1999; Halstead *et al.*, 2001), whereas severe dengue had been endemic to South-East Asia since the 1950s (Gubler, 1998a). Since the early 1980s, Asian DENV-2 has largely replaced the American subtype throughout the Caribbean (Lewis *et al.*, 1993; Vorndam *et al.*, 1994; Foster *et al.*, 2003). In light of this dynamic epidemiological history and evidence associating genotype with virulence (Gubler *et al.*, 1978; Rico-Hesse, 1990; Leitmeyer *et al.*, 1999; Cologna & Rico-Hesse, 2003) and/or number of infections (Bennett *et al.*, 2003), a detailed analysis of molecular evolution in PR

following the DENV-2 subtype replacement is essential to investigate the possible link between viral evolution and disease incidence.

In this study, we apply a longitudinal phylogenetic approach to a 20 year sequence dataset to examine patterns of molecular evolution and compare them with epidemiological observations. We focus on three genes that are critical to virus structure and host-cell entry (capsid, membrane and envelope) and that together represent over 20% of the approximately 11 kbp DENV-2 genome. Our results document evolutionary patterns in an emergent virus and contrast the relative roles of natural selection and genetic drift on a dengue serotype since it became re-established in densely populated PR.

## METHODS

We sequenced the three structural genes – capsid, membrane and envelope – for 91 DENV-2 isolates obtained from PR and other parts of the Americas since the early 1980s. Isolates were sampled from the US Centers for Disease Control and Prevention (CDC) serum bank; those from PR were selected from years with differing epidemic profiles (Fig. 1). Sample years with relatively high incidences of DENV-2 include 1988 ( $n=13$  isolates) and 1994 ( $n=19$ ). Sample years with relatively low incidences of DENV-2 include 1986 ( $n=2$ ), 1991 ( $n=16$ ), 1997 ( $n=1$ ), 1998 ( $n=14$ ) and 2001 ( $n=9$ ). In addition to these 74 PR isolates, we sequenced a virus isolated from Jamaica in 1983, the earliest available representative of the



**Fig. 1.** Prevalence of dengue virus in PR since 1981. Years included in this study are marked on the x axis with a black bar; the asterisk denotes the samples from Jamaica. Year labels begin in January and include all isolates collected until the January of the following year (next label). Of the dengue cases reported annually (solid black line, right axis), a subset is submitted to the CDC, isolated and identified by serotype (shaded areas, left axis, plotted against month/year of isolation). The proportions that were DENV-2 are shaded lighter grey. Because of dengue's variable aetiology, it often goes unreported, and thus the number of recorded cases underrepresents the true number of dengue infections by up to an estimated factor of 50–100 (WHO, 1999).

replacement subtype in the region (Deubel *et al.*, 1986, 1988). To provide a regional context for the PR data, we sequenced 16 additional isolates from various dates and locations throughout the Caribbean basin and Central and South America. Isolate label, GenBank accession number (Benson *et al.*, 2004), location and year of isolation, as well as the dataset in which they are included for analysis, can be found in Supplementary Table S1 (available in JGV Online). Isolates were selected randomly without regard to symptoms. All samples sequenced in this study had low-passage histories (maximum of two), reducing the risk of *in vitro* artificial selection. RNA extractions were made from the first C6/36 *Aedes albopictus* cell-culture passage or from the first *Toxorhynchites amboinensis* mosquito passage. To further eliminate potential biases due to artificial selection, samples were not processed in temporal (year) order.

We extracted sample RNA by using QIAamp Viral RNA Mini kits (Qiagen), purified the extraction with QIAquick PCR purification kits (Qiagen) and amplified each isolate for capsid, membrane and envelope genes (2322 bp of the 11 kbp viral genome or 21 %) by using one-step RT-PCR according to manufacturer's protocols (RT-PCR conditions and primer sequences can be obtained from the corresponding author). We sequenced both strands of the amplified products by using forward and reverse primers in standard dye-labelling reactions. Sequence data were collected on an ABI 377 slab-gel automated sequencer (Applied Biosystems) and MegaBACE 1000 96-capillary sequencer, edited, compiled and eventually aligned in SEQUENCHER 4.2 (Gene Codes Corporation). We imported aligned sequences into PAUP\* (Swofford, 2002) for phylogenetic analysis.

In addition to the 91 DENV-2 sequences obtained, several previously published sequences were incorporated into the alignments for phylogenetic analyses. We generated three phylogenetic trees to address (i) the evolutionary relationships of DENV-2 strains circulating in PR relative to known DENV-2 subtypes, (ii) the pattern of genetic change within PR and related regions and (iii) the influence of positive selection on this change, referred to as 'DENV-2 Global', 'DENV-2 SubtypeIIIb' and 'DENV-2 SubtypeIIIb Selection' datasets, respectively. The DENV-2 Global dataset combines a subset ( $n=31$ ) of the PR and other American isolates that we obtained, with 29 envelope and full structural-gene sequences (when available) for all DENV-2 subtypes, as well as a sequence from sylvatic DENV-2 (MAL P8.140770 from Malaysia, 1970; GenBank no. AF231717; Wang *et al.*, 2000) as an outgroup [Fig. 2a; Supplementary Table S1 (available in JGV Online)]. The DENV-2 SubtypeIIIb dataset includes 82 PR and American isolates in addition to 19 published sequences of the same subtype, including the earliest representative in the region, Jamaica 1983 (GenBank no. M20558; Deubel *et al.*, 1986, 1988) and the closest global representative, China 1985 (GenBank no. AF119661) as outgroup (Fig. 2b; Supplementary Table S1, available in JGV Online). The DENV-2 SubtypeIIIb Selection dataset includes only our PR and American isolates of the subtype with their complete capsid-, membrane- and envelope-gene sequences (83 isolates, including the outgroup; phylogeny not shown; see Supplementary Table S1, available in JGV Online). Maximum-likelihood (ML) phylogenetic trees were estimated by using the GTR +  $\Gamma$  + I model of nucleotide substitution, with the GTR substitution matrix, base composition, gamma distribution of among-site rate variation ( $\Gamma$ ) and proportion of invariant sites (I) all estimated from the data. The only exception was capsid for the DENV-2 SubtypeIIIb Selection dataset, whose best-fit model of evolution included two rather than six substitution rates (HKY85 + I +  $\Gamma$ ; Hasegawa *et al.*, 1985). Parameter values are available from the corresponding author upon request. To assess support for the ML tree topologies, we used Bayesian MCMC (Metropolis-Hastings Markov chain Monte Carlo) tree-sampling methods to generate posterior probabilities for each node, allowing substitution rate to vary by codon position, sampling four chains of  $1 \times 10^6$ – $2 \times 10^6$  generations every 100 generations with a burn-in of 2000–4000. Chain length was determined

based on convergence of likelihood values, giving effective sample sizes of over 400 (implemented in MrBayes v. 3; Huelsenbeck *et al.*, 2001; Ronquist & Huelsenbeck, 2003). Because this latter method is based on ML, it is consistent with our analytical approach and is the preferred method. However, standard bootstrap-support values, based on 1000 replicate neighbour-joining (NJ) trees under the ML substitution model described above, were also included for comparison.

To test for recombination among the DENV-2 isolates sequenced, we used the SplitsTree 3.2 package (Huson, 1998), which employs split decomposition to identify conflicts in phylogenetic signal (such as those caused by recombination), and other previously described methods (Bennett *et al.*, 2003). We also assessed the overall degree of recombination in the datasets by using a parsimony-informative sites ML test (PIST; Worobey, 2001).

Rates of nucleotide substitution were estimated by using an ML method available in the TipDate program (Rambaut, 2000), which compares the branch lengths of viral sequences sampled at different times. We compared the following models of substitution rate by using likelihood-ratio tests: different rates for every branch, DR; a single rate with dated tips, SRDT; and a single rate that varied linearly through time, VRDT (Rambaut, 2000).

To assess the extent of adaptive evolution in DENV-2 in PR, we compared rates of non-synonymous ( $d_N$ ) versus synonymous ( $d_S$ ) substitutions per site for each of the three genes sequenced in this study. Sites with disproportionately high relative rates of  $d_N$  across phylogenetic history provide strong evidence for positive selection (Yang *et al.*, 2000). We used a maximum-likelihood approach to compare models of evolution that allow  $d_N/d_S$  to vary across sites based on an ML tree for the DENV-2 SubtypeIIIb Selection dataset (Yang *et al.*, 2000). One model specifies a distribution of  $d_N/d_S$  classes across sites that are constrained to be  $\leq 1.0$  (model M7), thereby specifying neutral evolution, whereas the more complex M8 model incorporates an additional class of codons where  $d_N/d_S$  can be  $> 1$ , thus allowing for positive selection. These models were compared by using standard likelihood-ratio tests. To identify individual codons probably subject to positive selection, we applied a Bayesian approach to generate posterior probabilities of a given  $d_N/d_S$  class for each amino acid site, such that sites with high probabilities ( $> 0.99$ ) of falling into  $d_N/d_S$  category  $> 1$  are most likely to have been under positive selection. We also tested for selection on individual viral lineages by comparing the M0 model, in which each branch is assumed to have the same  $d_N/d_S$  ratio, with the FR ('free-ratio') model in which each branch is allowed to have a different  $d_N/d_S$  ratio (Yang *et al.*, 2000). All of these analyses were performed by using CODEML from the PAML package (Yang, 1997).

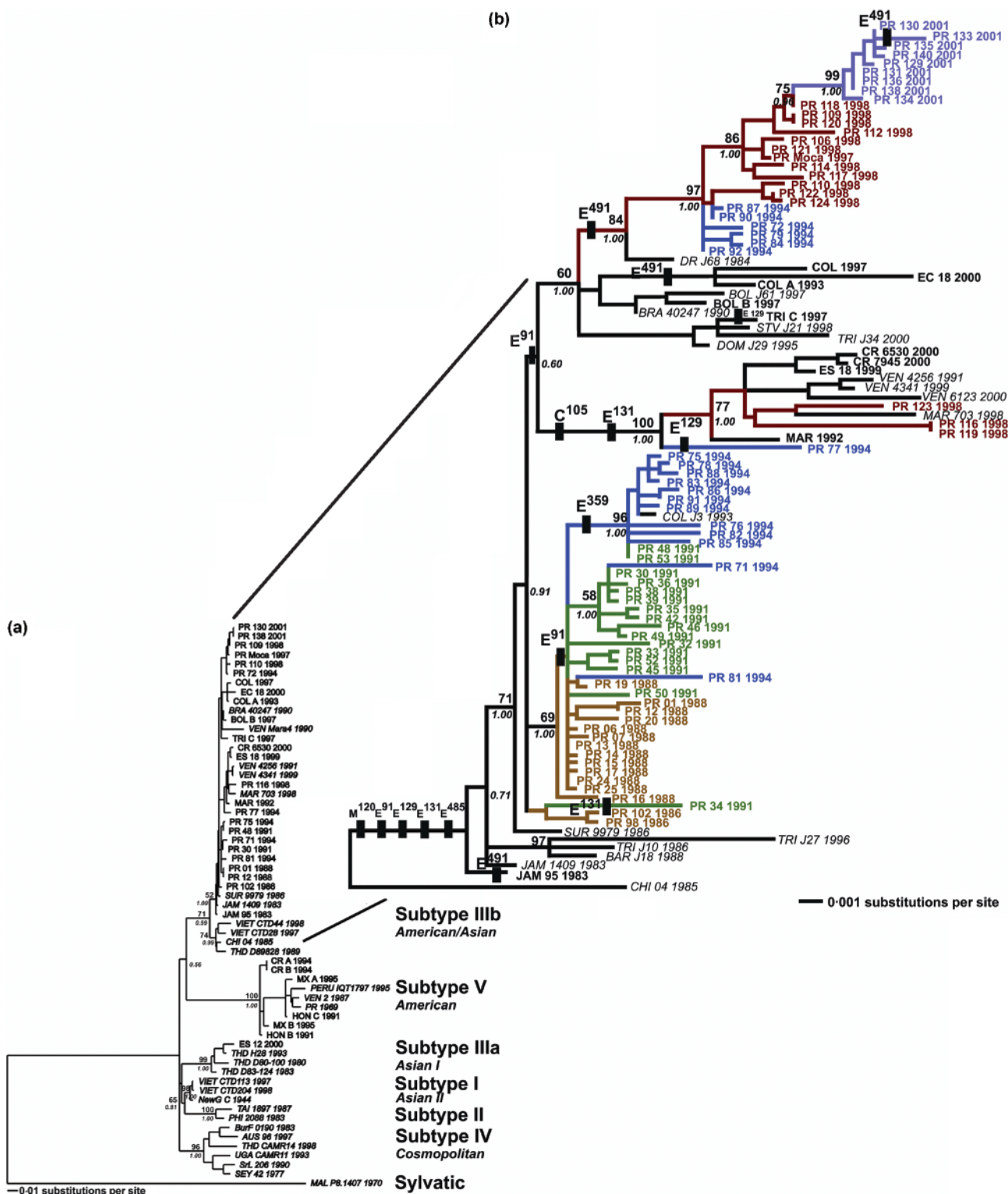
## RESULTS

DENV-2 in PR has exhibited a dynamic epidemiological history, with continuous transmission beginning in 1986 (Fig. 1). Since then, there have been seasonal and annual fluctuations in the number of viruses isolated by the CDC dengue-monitoring programme. Annual fluctuations ranged between low and high years, differing by over ninefold and peaking in 1994 (Fig. 1). This was the largest dengue outbreak in PR's history, with record numbers of hospitalizations ( $n=2004$ ), DHF cases ( $n=139$ ) and deaths ( $n=40$ ) (data not shown; Rigau-Pérez *et al.*, 2001).

Ninety of the 91 DENV-2 viruses collected between 1983 and 2001 from the Americas fell into two distinct subtypes: V, the American subtype, and IIIb, the Asian/American subtype (Fig. 2a). The single remaining isolate (El Salvador

2000) was most similar to isolates collected from Thailand between 1980 and 1993 [denoted subtype IIIa by Lewis *et al.* (1993) or Asian genotype I by Twiddy *et al.* (2002a)] (Fig. 2a). Caribbean isolates collected after 1981 were all of subtype IIIb. Throughout the study area, subtype V was not found after 1995. Subtypes IIIb and V differ at approximately 7.0%

of their nucleotide sequence (Fig. 2a). ML estimates for the number of non-synonymous and synonymous substitutions along the branch leading to subtype V were 8.1 and 153, respectively ( $d_N/d_S=0.0222$ ). The equivalent estimates for the branch leading to subtype IIIb were 1.9 and 46.3, respectively ( $d_N/d_S=0.0177$ ) (Fig. 2a). Although the



bootstrap support for the subtype IIIb clade based on the genes that we have examined is not compelling (71%), this group has been historically well recognized (Lewis *et al.*, 1993; Twiddy *et al.*, 2002a) and its cohesiveness is supported by a high ML-based Bayesian posterior node probability (0.99).

Within the American subtype IIIb, Asian and American variants fell into two groups, the first supported by 0.99 Bayesian posterior probability or in 74% of bootstrap replicates and the second by 1.00 Bayesian posterior probability or in 52% of bootstrap replicates (Fig. 2a). ML estimates for the number of non-synonymous versus synonymous substitutions along their branches were 3.8 and 9.4, respectively ( $d_N/d_S = 0.1704$ ), leading to the American lineage and 1.4 and 13.1, respectively ( $d_N/d_S = 0.0458$ ), leading to the Asian lineage. Most of the amino acid substitutions distinguishing these two lineages were conservative, but did include replacement of glutamine with leucine (L) at envelope site (E-) 131, replacing a hydrophilic with a hydrophobic residue. The closest Asian subtype IIIb virus related to the lineage appearing in Jamaica in 1983 was isolated in China in 1985 and was used as a root to the American subtype IIIb phylogeny in Fig. 2(b).

Sequence differences within the American subtype IIIb accumulated throughout our sampling period. We found no evidence that recombination had been important during this evolution based on either graphical split-tree decomposition or tests against expected levels of homoplasy under clonal evolution (PIST,  $P < 0.6$ ). Analysis of the pattern of nucleotide substitutions suggested that evolutionary rates varied branch by branch in our best-supported phylogenetic tree (Fig. 2b). Specifically, the DR model was a better fit than the simpler SRDT model ( $P < 0.01$ ), thereby rejecting the molecular clock. However, most of the rate variation could be attributed to substitutions within one lineage containing both temporally and geographically diverse DENV-2 isolates

collected from the Americas. This lineage included isolates from Martinique (1992, 1998), PR (1994, 1998), Costa Rica (1999, 2000) and Venezuela (1999, 2000). Removing these isolates from rate estimates provided a much better fit to an SRDT model ( $P = 0.2$ ), giving us a substitution-rate estimate of  $8.34 \times 10^{-4}$  [ $5.97 \times 10^{-4}$ ,  $10.15 \times 10^{-4}$  (mean  $\pm$  95% ML confidence interval)] substitutions per site year<sup>-1</sup>. The age of the most recent common ancestor (MRCA) for this PR subset is 20.42 years [19.72, 22.26 (mean  $\pm$  95% ML confidence interval)].

Within PR, there is evidence for *in situ* time-ordered evolution punctuated by significant evolutionary shifts in the predominant circulating lineage (Fig. 2b). For example, DENV isolates collected from 1988 to 1991 fall within a distinct clade, with individuals grouped clearly according to time of isolation (Fig. 2b). However, the large 1994 outbreak is marked by the presence of three distinctive lineages that are more or less equally divergent from each other. Although one group is nested within the lineage present in 1991, the two remaining groups fall into separate clades (Fig. 2b), both of which contain isolates collected from regions outside PR. One of these lineages apparently persisted at low levels in PR from 1994 until 1998, whereas the other became the dominant lineage and showed a similarly time-ordered phylogenetic pattern. This lineage may have arisen from a Dominican Republic isolate from 1984 (Fig. 2b). In addition, several other non-PR isolates from the Caribbean and South America between 1990 and 2000 are also associated with this lineage, further supporting its foreign origin.

Most significantly, the fixation of variation in DENV-2 subtype IIIb from PR appears to have been shaped, in part, by adaptive evolution. Positive selection was detected by measuring rates of non-synonymous ( $d_N$ ) and synonymous ( $d_S$ ) nucleotide substitutions per site across the subtype IIIb

**Fig. 2.** (a) Maximum-likelihood (ML) consensus tree showing the phylogenetic relationships between DENV-2 subtypes based on structural (capsid, membrane and envelope) gene sequences (when available, up to 2322 bp), including some previously published data (tip labels in italics). Samples are coded by location, number and year of isolation: Australia (AUS), Bolivia (BOL), Brazil (BRA), Burkino Faso (BurF), China (CHI), Colombia (COL), Costa Rica (CR), Ecuador (EC), El Salvador (ES), Honduras (HON), Jamaica (JAM), Malaysia (MAL, sylvatic), Martinique (MAR), Mexico (MEX), New Guinea (NewG C), Peru, Philippines (PHI), Puerto Rico, (PR), Seychelles (SEY), Sri Lanka (SrL), Surinam (SUR), Taiwan (TAI), Thailand (THD), Trinidad (TRI), Uganda (UGA), Venezuela (VEN) and Vietnam (VIET). Subtypes are labelled by number [in bold, from Lewis *et al.* (1993)] and distribution [in italics, from Twiddy *et al.* (2002a)]. Consensus tree was generated from Bayesian MCMC methods, four chains of length  $10^6$ , sample frequency 100, with a 2000 tree burn-in. Bootstrap-support values of 1000 NJ replicates using the ML model of evolution and Bayesian posterior probabilities appear above and below each node, respectively. (b) Subtype IIIb ML tree based on structural-gene sequences for 74 isolates of DENV-2 from PR along with two Jamaican isolates (D2.95 and JAM 1983, the first associated cases in the region) and 25 other foreign isolates from Barbados (BAR), Dominica (DOM), Dominican Republic (DR) and St Vincent (STV), in addition to several of those listed above. Branches are colour-coded by year of sample isolation for PR sequences; foreign isolates are in black. Sample labels are italicized if published previously (see Supplementary Table S1, available in JGV Online, for GenBank accession numbers). Bootstrap-support values, based on 1000 NJ replicates under the ML model of nucleotide evolution, and Bayesian posterior probabilities appear above and below nodes, respectively. Unambiguous amino acid changes in capsid (C), membrane (M) and envelope (E) genes are indicated along the major lineages (changes at tips, including those specific to the China outgroup but not the Asian IIIb lineage, were omitted), marked with black bars and the amino acid position within their respective genes.

PR phylogeny (DENV-2 SubtypeIIIb Selection dataset). Although our analysis of the capsid and membrane genes suggested that their evolution is shaped by negative selection ( $d_N/d_S \ll 1$ ), we found statistically significant evidence ( $P=0.02$ ) for positive selection ( $d_N/d_S=3.2$ ) acting at a small number of sites in the envelope gene. By using a Bayesian approach, this selection was assigned to four amino acid sites within the DENV-2 envelope gene – positions 91, 129, 131 and 491 (all  $P>0.98$ ), all of which change frequently within the phylogeny and often define major clades (Fig. 2b). It should also be noted that these selection analyses are based on codon variation and are therefore unable to detect a single selection event that results in fixation of a mutation, in which all descendents become monomorphic.

## DISCUSSION

We have performed a comprehensive analysis of DENV-2 evolution in PR during the period following the introduction of epidemic DHF into the region (Kouri *et al.*, 1989; Guzman *et al.*, 1984). Our results show that subtype IIIb has been the dominant lineage in PR and parts of the Americas for the past 10–20 years. Subtype IIIb has been distributed extensively throughout the Caribbean (PR, Martinique, Trinidad, Jamaica) and South and Central America (Columbia, Costa Rica, El Salvador) and, in those localities from which we have isolates of both subtypes (i.e. Costa Rica, PR and Venezuela), subtype IIIb has been isolated most recently. Subtype V was last isolated in the Caribbean in 1981 and overall in 1995 (Rico-Hesse, 1990; Guzman *et al.*, 1995; Leitmeyer *et al.*, 1999; Sariol *et al.*, 1999; Foster *et al.*, 2004). This and our extensive temporal sampling in PR since that time, which has yielded only subtype IIIb (Fig. 2a), supports the observation that subtype IIIb is replacing the original American subtype V in the Caribbean and much of South and Central America (Foster *et al.*, 2004).

Since DENV-2 subtype IIIb was established in 1986, much of the variation in PR has arisen *in situ*, as suggested by our phylogenetic analysis, particularly the observation that two major lineages, from 1988 to 1994 and from 1994 to 2001, are restricted to PR. As such, a Colombian isolate from 1993 that fell within the other PR 1994 clade is likely to represent an exportation of a PR virus to that locality (Fig. 2b). However, 1994, the year of PR's largest and most severe dengue epidemic (Rigau-Pérez *et al.*, 2001), was unique in that two novel lineages appeared in this country. Both lineages were divergent from the predominant lineage present in PR prior to 1994 and probably have foreign origins. The lineage that became dominant in PR after 1994 persisted until 2001 and was related to several foreign isolates from as early as 1984 from the Dominican Republic (GenBank no. AY484625; Foster *et al.*, 2004), further supporting a foreign source for this lineage. The third lineage present in 1994, consisting mainly of South and Central American isolates from 1992 to 2000, includes only four divergent PR isolates from 1994 and 1998, suggesting

that it never became established in PR. Overall, the complex relationships between DENV-2 strains from PR and other localities underline the importance of spatial heterogeneity and migration on dengue disease dynamics in island localities.

The spread of subtype IIIb throughout the Caribbean basin has been relatively fast, although the exact date of its introduction into the Americas remains problematic. An isolate from the 1981 Cuban epidemic that marked the onset of epidemic DHF in the region was not subtype IIIb, but similar to the New Guinea C 1944 reference strain (GenBank no. M29095; Guzman *et al.*, 1995; Sariol *et al.*, 1999). This is clearly an issue that needs to be investigated further. However, our rate estimations suggest that subtype IIIb in PR originated only 20 years before the most recent sample analysed, i.e. in 1981, and the subtype was definitely in Jamaica by 1983 (Foster *et al.*, 2004). The swift establishment and spread of subtype IIIb since 1983 into at least 12 American countries, where it has seemingly displaced subtype V in the Caribbean basin, has also been accompanied by relatively few amino acid changes that may nevertheless have had dramatic effect on epidemic potential.

A key question in dengue research is, therefore, why DENV-2 subtype IIIb has spread, and in some places displaced subtype V, in the Americas? This replacement event may simply reflect stochastic processes, such as changes in vector density (Kuno, 1997). Alternatively, its rapidity and association with increases in DHF and DSS cases may reflect intrinsic differences in viral transmissibility between the two subtypes, as suggested by recent experimental analyses (Cologna & Rico-Hesse, 2003). Moreover, amino acid differences between subtypes V and IIIb have been correlated with changes in virulence, with those in the envelope gene probably affecting host cell-binding efficiency (Sánchez & Ruiz, 1996; Kinney *et al.*, 1997; Leitmeyer *et al.*, 1999). More importantly, although most amino acid sites in subtype IIIb are conserved, indicating that they are subject to purifying selection, we found strong evidence for positive selection at four amino acid positions within the envelope gene – positions 91, 129, 131 and 431. The fact that these sites change multiple times across our phylogeny and often distinguish major monophyletic groups supports the idea that they are of great functional relevance.

The envelope gene encodes proteins on the viral surface that enable host-cell binding and entry via membrane fusion and also provide the primary target for host immune responses (Roehrig, 1997). Experiments verify that viral gene regions that interact specifically with host cells are evolutionarily constrained, particularly in viruses that infect multiple different hosts (Weaver *et al.*, 1999). Although accounting for a very small proportion of sites, the four amino acid sites under positive selection fall into immunogenic regions (reviewed by Roehrig, 1997). E-91, E-129 and E-131 are in the protein domain (II) responsible for host-cell membrane fusion, where E-91 is just seven residues downstream of the highly conserved (in flaviviruses) fusion loop (transmembrane

region) and *E*-129 and *E*-131 are in the 'elongated, finger-like' section of domain II, which becomes folded into domain III during membrane fusion (Modis *et al.*, 2004). Although substitutions at *E*-91 and *E*-129 were all conservative (aliphatic, hydrophobic residues before and after), *E*-131 has undergone non-conservative substitutions, from an aliphatic, hydrophobic residue to a moderately hydrophilic polar amide. Selected site *E*-491 occurs in the 53 residue 'stem' portion (connecting domain III to the viral transmembrane anchor) of the envelope protein, the section that becomes folded with domain II during membrane fusion (Modis *et al.*, 2004). Substitutions at *E*-491 were also relatively conservative (again amongst aliphatic, hydrophobic residues). Another site, *E*-359, distinguishes the 1994 clade originating in PR 1988 (Fig. 2b). Although this latter site underwent a mutation only once in subtype IIIb's evolutionary history in PR, it was a non-conservative amino acid substitution (from threonine, T, to alanine, A) resulting in the substitution of an amino acid potentially involved in phosphorylation/glycosylation (T) with one that is not (A). The overall pattern of amino acid substitution suggests that they may be changing in concert: changes in *E* domain I (*E*-91, *E*-129 and *E*-131) were often accompanied by changes in *E* domain III (*E*-491 or *E*-359) (Fig. 2b), although never in positions known to interact directly (Modis *et al.*, 2004). Although we lack the power (in terms of number of replications of amino acid-defined clades) to test the significance of this pattern, it is not unlikely given the physical inter-lacing of domains I and III during membrane fusion (Modis *et al.*, 2004). Previous studies also report sporadic positive selection in envelope amongst some DENV-2 subtypes (Twiddy *et al.*, 2002a, b), although purifying selection is clearly the dominant evolutionary pressure acting on DENV. Together, these results suggest that positive selection may be acting on host-cell infection rates or immune evasion, which could ultimately affect viral reproduction. Substitutions at some of the selected sites also define phylogenetic structure, indicating their proliferation in descendent DENV populations: synapomorphies at *E*-91 and/or *E*-491 distinguish the 1988–1994 and 1994–2001 PR clades, both high years of DENV-2 prevalence in PR.

In DENV, associations between genotypic and epidemiological change are particularly important for inferring changes in virus phenotype, as experimental systems are lacking. The peak years of DENV-2 subtype IIIb transmission in PR were 1988 and 1994 (Fig. 1). The predominant viruses in 1988 are distinguished by a conservative amino acid substitution in envelope and, albeit a positively selected site in an important region, the phenotypic implications are difficult to infer. Alternatively, the outbreak of 1988 may be due to the elevated numbers of immunologically naïve hosts, as this was the first major year of DENV-2 transmission (Fig. 2b). In contrast, 1994 consists of two distinct but equally prevalent clades, one derived from the 1988 group that became extinct after 1994, but was distinguished by a highly non-conservative amino acid substitution, and the other distinguished by a conservative substitution at a positively selected

site, which was the lineage that persisted until 2001 and had geographically widespread origins (Fig. 2b).

Although the foreign (1994) clade appears to have the greater epidemic potential based on its spatial and temporal distribution, it is currently difficult to associate specific phenotypic changes, including those at the positively selected sites, with both outbreak severity and clade dynamics. In particular, mutational changes defining these key outbreak clades were either relatively conservative (i.e. the foreign 1994 lineage) or counter to clade persistence (i.e. the native 1994 lineage). It is possible that the native 1994 clade underwent phenotypically relevant mutations to severe outbreak conditions, but that these variants did not persist in PR beyond 1994. This clade's extinction may be associated with these changes or a largely random event. Alternatively, and similar to the 1988 outbreak, the antigenic novelty of the foreign 1994 lineage relative to the native 1994 clade may account for its long-term success and a concomitant increase in numbers and severity of cases. Future studies should therefore endeavour to characterize the fitness values of individual amino acid changes and be coupled with molecular-epidemiological studies that aim to identify potentially important mutations or genomic regions that may drive disease dynamics.

## ACKNOWLEDGEMENTS

We would like to thank NIH-NAIAD (to W. O. M.) and the Wellcome Trust (to E. C. H.) for financial support. In addition, the RCMi and Score/BRIN programs at the University of Puerto Rico – Rio Piedras provided shared sequencing facilities and infrastructural support, as did the COBRE (Yanagihara) and INBRE (Smith) programmes at the University of Hawaii. Thanks also go to Durrell D. Kapan for valuable editorial comments.

## REFERENCES

- 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.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. (2004). GenBank: update. *Nucleic Acids Res* 32, D23–D26.
- Cologna, R. & Rico-Hesse, R. (2003). American genotype structures decrease dengue virus output from human monocytes and dendritic cells. *J Virol* 77, 3929–3938.
- Deubel, V., Kinney, R. M. & Trent, D. W. (1986). Nucleotide sequence and deduced amino acid sequence of the structural proteins of dengue type 2 virus, Jamaica genotype. *Virology* 155, 365–377.
- Deubel, V., Kinney, R. M. & Trent, D. W. (1988). Nucleotide sequence and deduced amino acid sequence of the nonstructural proteins of dengue type 2 virus, Jamaica genotype: comparative analysis of the full-length genome. *Virology* 165, 234–244.
- Dietz, V., Gubler, D. J., Ortiz, S., Kuno, G., Casta-Velez, A., Sather, G. E., Gomez, I. & Vergne, E. (1996). The 1986 dengue and dengue hemorrhagic fever epidemic in Puerto Rico: epidemiologic and clinical observations. *P R Health Sci J* 15, 201–210.

- Foster, J. E., Bennett, S. N., Vaughan, H., Vorndam, V., McMillan, W. O. & Carrington, C. V. F. (2003). Molecular evolution and phylogeny of dengue type 4 virus in the Caribbean. *Virology* **306**, 126–134.
- Foster, J. E., Bennett, S. N., Carrington, C. V. F., Vaughan, H. & McMillan, W. O. (2004). Phylogeography and molecular evolution of dengue type 2 virus in the Caribbean basin, 1981–2000. *Virology* **324**, 48–59.
- Gubler, D. J. (1993). Dengue and dengue hemorrhagic fever in the Americas. In *Dengue Hemorrhagic Fever*, WHO regional publication SEARO no. 22, pp. 9–22. Edited by P. Thoncharoen. New Delhi, India: WHO.
- 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. (1998a). Dengue and dengue hemorrhagic fever. *Clin Microbiol Rev* **11**, 480–496.
- Gubler, D. J. (1998b). The global pandemic of dengue/dengue haemorrhagic fever: current status and prospects for the future. *Ann Acad Med Singapore* **27**, 227–234.
- Gubler, D. J. (2002). Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. *Trends Microbiol* **10**, 100–103.
- Gubler, D. J., Reed, D., Rosen, L. & Hitchcock, J. R., Jr (1978). Epidemiologic, clinical, and virologic observations on dengue in the Kingdom of Tonga. *Am J Trop Med Hyg* **27**, 581–589.
- Guzman, M. G., Kouri, G., Morier, L., Soler, M. & Fernandez, A. (1984). A study of fatal hemorrhagic dengue cases in Cuba, 1981. *Bull Pan Am Health Organ* **18**, 213–220.
- Guzman, M. G., Deubel, V., Pelegrino, J. L., Rosario, D., Marrero, M., Sariol, C. & Kouri, G. (1995). Partial nucleotide and amino acid sequences of the envelope and the envelope/nonstructural protein-1 gene junction of four dengue-2 virus strains isolated during the 1981 Cuban epidemic. *Am J Trop Med Hyg* **52**, 241–246.
- Halstead, S. B. (1988). Pathogenesis of dengue: challenges to molecular biology. *Science* **239**, 476–481.
- Halstead, S. B., Streit, T. G., Lafontant, J. G., Putvatana, R., Russel, K., Sun, W., Kanesa-Thanan, N., Hayes, C. G. & Watts, D. M. (2001). Haiti: absence of dengue hemorrhagic fever despite hyperendemic dengue virus transmission. *Am J Trop Med Hyg* **65**, 180–183.
- Hasegawa, M., Kishino, H. & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* **22**, 160–174.
- Huelsenbeck, J. P., Ronquist, F., Nielsen, R. & Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**, 2310–2314.
- Huson, D. H. (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68–73.
- Kinney, R. M., Butrapet, S., Chang, G.-J. J., Tsuchiya, K. R., Roehrig, J. T., Bhamarapavati, N. & Gubler, D. J. (1997). Construction of infectious cDNA clones for dengue 2 virus: strain 16681 and its attenuated vaccine derivative, strain PDK-53. *Virology* **230**, 300–308.
- Kliks, S. C., Nisalak, A., Brandt, W. E., Wahl, L. & Burke, D. S. (1989). Antibody-dependent enhancement of dengue virus growth in human monocytes as a risk factor for dengue hemorrhagic fever. *Am J Trop Med Hyg* **40**, 444–451.
- Kouri, G. P., Guzman, M. G., Bravo, J. R. & Triana, C. (1989). Dengue haemorrhagic fever/dengue shock syndrome: lessons from the Cuban epidemic, 1981. *Bull W H O* **67**, 375–380.
- Kuno, G. (1997). Factors influencing the transmission of dengue viruses. In *Dengue and Dengue Hemorrhagic Fever*, pp. 61–88. Edited by D. J. Gubler & G. Kuno. London: CAB International.
- Leitmeyer, K. C., Vaughn, D. W., Watts, D. M., Salas, R., Villalobos de Chacon, I., Ramos, C. & Rico-Hesse, R. (1999). Dengue virus structural differences that correlate with pathogenesis. *J Virol* **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.
- Modis, Y., Ogata, S., Clements, D. & Harrison, S. C. (2004). Structure of the dengue virus envelope protein after membrane fusion. *Nature* **427**, 313–319.
- Rambaut, A. (2000). Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* **16**, 395–399.
- 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.
- Rigau-Pérez, J. G., Vorndam, A. V. & Clark, G. G. (2001). The dengue and dengue hemorrhagic fever epidemic in Puerto Rico, 1994–1995. *Am J Trop Med Hyg* **64**, 67–74.
- Roehrig, J. T. (1997). Immunochemistry of dengue viruses. In *Dengue and Dengue Hemorrhagic Fever*, pp. 199–219. Edited by D. J. Gubler & G. Kuno. London: CAB International.
- Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.
- Sánchez, 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. *J Gen Virol* **77**, 2541–2545.
- Sariol, C. A., Pelegrino, J. L., Martinez, A., Arteaga, E., Kouri, G. & Guzman, M. G. (1999). Detection and genetic relationship of dengue virus sequences in seventeen-year-old paraffin-embedded samples from Cuba. *Am J Trop Med Hyg* **61**, 994–1000.
- Swofford, D. L. (2002). PAUP\*: Phylogenetic analysis using parsimony, version 4.0b. Sunderland, MA: Sinauer Associates.
- Thein, S., Aung, M. M., Shwe, T. N., Aye, M., Aung, Z., Aye, K., Aye, K. M. & Askov, J. (1997). Risk factors in dengue shock syndrome. *Am J Trop Med Hyg* **56**, 566–572.
- 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.
- Vorndam, V., Nogueira, R. M. R. & Trent, D. W. (1994). Restriction enzyme analysis of American region dengue viruses. *Arch Virol* **136**, 191–196.
- Wang, E., Ni, H., Xu, R., Barrett, A. D. T., Watowich, S. J., Gubler, D. J. & Weaver, S. C. (2000). Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. *J Virol* **74**, 3227–3234.
- 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 hemorrhagic fever. *Lancet* **354**, 1431–1434.

**Weaver, S. C., Brault, A. C., Kang, W. & Holland, J. J. (1999).** Genetic and fitness changes accompanying adaptation of an arbovirus to vertebrate and invertebrate cells. *J Virol* **73**, 4316–4326.

**WHO (1999).** *Strengthening Implementation of the Global Strategy for Dengue Fever/ Dengue Haemorrhagic Fever Prevention and Control*. Report of the Informal Consultation, 18–20 October 1999, WHO HQ, Geneva [WHO Report WHO/CDS/(DEN)/IC/2000.1]. <http://www.who.int/csr/resources/publications/dengue/whocdsdenic20001.pdf>

**Worobey, M. (2001).** A novel approach to detecting and measuring recombination: new insights into evolution in viruses, bacteria, and mitochondria. *Mol Biol Evol* **18**, 1425–1434.

**Yang, Z. (1997).** PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**, 555–556.

**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.