Virology **303,** 110–119 (2002)
doi:10.1006/viro.2002.1686

Diversity and Evolution of the Envelope Gene of Dengue Virus Type 1

Ana P. Goncalvez,*,† Ananias A. Escalante,* Flor H. Pujol,* Juan E. Ludert,* Duilia Tovar,‡ Rosa A. Salas,‡[1] and Ferdinando Liprandi*,[2]

*Instituto Venezolano de Investigaciones Científicas, Caracas, Venezuela; †Facultad de Ciencias de la Salud, Universidad de Carabobo, Valencia, Venezuela; and ‡Instituto Nacional de Higiene "Rafael Rangel", Caracas, Venezuela

Received December 7, 2001; returned to author for revision January 18, 2002; accepted July 22, 2002

The genetic diversity and phylogenetic relationships of a collection of strains of dengue virus type 1 (DV-1), isolated from different parts of the world, were investigated. Phylogenetic trees derived from the complete sequence of the E gene of 44 strains suggested the existence of five genetic types defined by a maximum nucleotide divergence within each group of 6%. The 22 strains from America were classified into a single genetic type that included strains associated either with classical dengue or hemorrhagic dengue episodes. Using a maximum likelihood procedure based on a single rate with dated tips model and substitution rates calculated at the third codon position, evolution of the five DV-1 genotypes was shown to conform to a molecular clock. The average rate of evolution was estimated to be approximately 16.2×10^{-4} substitutions/third codon position site/year. Using this estimate, divergence among the DV-1 genotypes was calculated to have occurred approximately 100 years ago. Very low average value of the ratio of nonsynonymous-to-synonymous nucleotide substitutions, relative to the respective sites (0.046), indicated that the evolution of the E gene of the DV-1 is subject mostly to purifying selection. © 2002 Elsevier Science (USA)

Key Words: dengue; dengue virus; molecular evolution; molecular epidemiology; genetic diversity; envelope; nucleotide substitutions; molecular clock.

INTRODUCTION

Dengue viruses (DV) (family *Flaviviridae,* genus *Flavivirus*) occur as four antigenically distinct serotypes, which do not induce cross protection. Infection with any of these leads to a mild, self-limiting febrile illness (dengue fever, DF). DF is currently the most important mosquito-transmitted viral disease: about 100 million cases of DF occur annually, with some 2 billion people at risk of infection in tropical and subtropical regions of Africa, Asia, and the Americas (Monath, 1994; Monath and Heinz, 1996). A more severe form of the disease, dengue hemorrhagic fever/dengue shock syndrome (DHF/DSS), is responsible for a high mortality rate, especially in children. A secondary infection by a heterologous DV serotype has been established as the main risk factor for the occurrence of DHF/DSS (Halstead, 1988). However, reports of primary infections resulting in DHF/DSS suggest that differences in virulence of DV strains may also be involved (Rosen, 1977). Also, the emergence of DHF/DSS has been associated with the introduction of a specific type of DV: the emergence of DHF in the American continent, first in Cuba in 1981 and later in other countries in 1989, has been associated with the introduction of a genetic type of DV serotype 2, which originated from Southeast Asia, with the virulence potential to produce DHF (Rico-Hesse, 1990; Rico-Hesse *et al.,* 1997). Conversely, the "native" American genotype has been associated thus far only with mild disease (DF) (Rico-Hesse *et al.,* 1997; Leitmeyer *et al.,* 1999; Watts *et al.,* 1999). Therefore characterization of intraserotypic genetic variation, in addition to allowing the pursuit of global distribution and geographic movements of virus strains (Lanciotti *et al.,* 1994, 1997), should also lead to the identification of genetic differences associated with clinical severity of human infection (Leitmeyer *et al.,* 1999). Genetic variation has been studied with the four dengue serotypes, the analysis being based mainly on the comparisons of sequences of genes coding for structural proteins (Rico-Hesse, 1990; Lewis *et al.,* 1993; Lanciotti *et al.,* 1994, 1997; Chungue *et al.,* 1995). Of the four serotypes, DV-1 has been the less extensively characterized in terms of genetic diversity. This has been addressed in two previous studies based on the analysis of either a fragment in the 5′ end of the E gene (Chungue *et al.,* 1995) or a 240-nucleotide fragment spanning the E-NS1 junction (Rico-Hesse, 1990). The existence of three to five different genotypes has been proposed, with one or two genotypes apparently circulating in the Americas. DV-1 was introduced in America in 1977, being responsible in the next few years for several epidemics in different countries (Pinheiro, 1989; Pinheiro and Cor-

[1] Present address: PHI Unit, Caribbean Epidemiology Center, 16–18 Jamaica Boulevard, Federation Park, Port of Spain, Trinidad, West Indies.

[2] To whom correspondence and reprint requests should be addressed at Laboratorio de Biología de Virus, CMBC, IVIC Apdo. 21827, Caracas 1020[a], Venezuela. Fax: 58-212-5041623. E-mail: fliprand@ivic.ve.

0042-6822/02 $35.00
© 2002 Elsevier Science (USA)
All rights reserved.

bet, 1997). Since then, it has been maintained mainly in an endemic pattern associated with classical DF and sporadically associated with DHF.

The objective of the present study was to revise the molecular epidemiology and evolution of DV-1 viruses based on the sequence analysis of the envelope gene (E) of a collection of DV-1 strains. Sequence variation within the E gene allowed the classification of DV-1 into five discrete genetic lineages, which evolution complied to with a molecular clock model. The estimated evolution rate indicated that divergence of different genetic lineages had occurred approximately 100 years ago.

RESULTS

Nucleotide sequence analysis

The nucleotide sequences of the E gene from 31 strains of DV-1 were determined. The strains sequenced included seven isolates from Venezuela spanning from 1994 to 1997 plus 24 isolates from different parts of the world. These were aligned with 13 sequences already reported to the GenBank (Table 1). Five additional sequences deposited in the GenBank (D00501 to D00505) (Chu *et al.,* 1989) were not included in the analysis, for reasons explained under Discussion. All the E gene sequences were 1485 nucleotides long, with the exception of the Singapore 90 strain (1482 nucleotides), which presented a deletion of three nucleotides at positions 1045 to 1047. Comparisons of the 44 DV-1 E gene sequences revealed base substitutions scattered throughout the entire gene, without a particular region of hypervariability. The majority of mutations occurred at the third codon position and was therefore silent. The nucleic acid sequence divergence among DV-1 strains ranged from 0.2 to 9.4%.

Phylogenetic analysis

Phylogenetic analysis of the 44 DV-1 E gene nucleotide sequences was performed using distance (neighbor-joining), parsimony, and maximum-likelihood methods. The three methods produced trees with a topology suggestive of the existence of five genetic types (Fig. 1): genotype I, represented by two early isolates from the 1940s, including the prototype strain Hawaii, and by strains from Southeast Asia, China, and East Africa; genotype II, represented by strains isolated in Thailand in the 1950s and the 1960s; genotype III, represented by one sylvatic strain from Malaysia; genotype IV, including isolates from West Pacific islands and Australia; genotype V, grouping all the strains isolated from America, three strains from West Africa, and two strains from Asia. Within each genotype, observed sequence differences ranged from 0.2 to 6.0% (Table 2), whereas among viruses of different genotypes, sequence divergence ranged from 6.0 to 9.4%. There were the following exceptions: strains Hawa45 and Japa43 (genotype I) presented sequence divergences lower than 6% (between 5.4 and 5.9%) with some strains of genotypes II, III, and IV. Similarly strain Mala72 (genotype III) presented divergences between 5.6 and 5.9% with some strains of genotypes I and II.

Although phylogenetic analysis using the three methods yielded trees with similar subdivisions into five lineages, the branching order of the basal nodes differed between methods, and with the ML analysis varied according to the outgroup used and the model of nucleotide substitutions employed, whether with constant or gamma-distributed rates. In trees rooted using as outgroup representative strains of either dengue 3 or each serotype 2, 3, and 4 (Fig. 1A), the basal nodes could not be resolved since they were unsupported by significant bootstrap values. In a condensed tree with the latter outgroup, a multifurcating origin of the five clades was supported by a 53% bootstrap value. Instead in unrooted trees the topology of key nodes was supported by high (>70%) bootstraps values for all genotypes, except for the sylvatic strain P72–1244 (Mala1972) (Fig. 1B). The topology of this strain is an important one, since this type of strain has been proposed as putative progenitor of the endemic–epidemic lineages (Gubler, 1997; Wang *et al.,* 2000). Consistent with this, the strain occupied a basal position in the E gene phylogeny based on a parsimony method (data not shown), as previously observed by Wang *et al.* (2000), but did not so in the phylogenies by neighbor-joining and maximum likelihood methods (Fig. 1). Phylogenetic comparisons between the two types of strains based on other regions of the DV-1 genome will be required to resolve the issue.

The sequences of the 22 American isolates, with a nucleotide identity greater than 96.7%, fell into a single group together with three African strains and two Asian strains (Burm76 and Sing90). American strains included representatives from 10 different countries spanning a 20-year period, from the 1977 introduction to the 1997 Venezuelan DV-1 epidemic (Salas *et al.,* 1998). The Venezuelan strains, which were all closely related, included five strains associated with classical dengue fever and two strains associated with hemorrhagic dengue fever (Table 1). Thus a single genotype of DV-1, with the potential to produce DHF, has been circulating in America since 1977.

Angola is the geographic origin proposed for the RIO H 36589 strain, since it was isolated from a Brazilian traveler who presented FD 1 week after his return to Brazil from Angola in 1988 (Tesh, personal communication). Nevertheless, this strain was more closely related to the American (96.7 to 99.1% similarity) than to the African strains (94.8 to 95.8%) (Fig. 1), suggesting its American origin.

TABLE 1

Dengue 1 Virus Isolates Used for Sequence Analysis

Strain	Code	Year[a]	Location	Passage history[b]	Accession No.[c]
125239	VeTac94	1994	Venezuela[d]	C6/36 1	AF425637
141602	VePor94	1994	Venezuela[e]	C6/36 1	AF425636
150172	VeDtF95	1995	Venezuela[e]	C6/36 1	AF425633
5345	VeMir95	1995	Venezuela[e]	C6/36 1	AF425635
5736	VeZul95	1995	Venezuela[e]	C6/36 1	AF425638
6222	VeAma95	1995	Venezuela[e]	C6/36 1	AF425632
28164	VeGua97	1997	Venezuela[d]	C6/36 1	AF425634
CAREC 778156	Gren77	1977	Grenada	C6/36 2	AF425618
PRS 228690	Jama77	1977	Jamaica	Mosq. ?, C6/36 2	AF425621
CAREC 780572	Trin78	1978	Trinid and Tobago	C6/36 2	AF425631
CAREC 86471	Trin86	1986	Trinid and Tobago	C6/36 2	AF425639
1298/TVP 951	Mexi80	1980	Mexico	Mosq. 2, C6/36 2	AF425623
1378	Mexi83	1983	Mexico[e]	Mosq. 2, C6/36 2	AF425624
495-1	Arub85	1985	Aruba[d]	C6/36 2	AF425609
BE AR 404147	Braz82	1982	Brazil	C6/36 2	AF425613
BeH 584526	Braz97	1997	Brazil	C6/36 3	AF425614
INS 347869	Colo85	1985	Colombia	C6/36 4	AF425616
INS 371869	Colo96	1996	Colombia	C6/36 2	AF425617
DEI 0151	Peru91	1991	Peru	C6/36 2	AF425626
Hawaii	Hawa45	1945	Hawaii	Monkey 1, Mosq. 6, C6/36 3	AF425619
AUS HCS1	Aust83i	1983	Australia	C6/36 2	AF425611
AUS H TI 7	Aust83ii	1983	Australia	C6/36 6	AF425612
P72-1244	Mal72	1972	Malaysia	?, C6/36 2	AF425622
PRS 228682	Phil74	1974	Philippines	C6/36 2	AF425627
PRS 228686	Burm76	1976	Burma	Mosq. ?, C6/36 2	AF425615
2543-63	Thai63	1963	Thailand	SM 5, C6/36 2	AF425629
PUO 359	Thai80	1980	Thailand	C6/36 2	AF425630
765101	Taiw87	1987	Taiwan	C6/36 5	AF425628
IBH 28328	Nige68	1968	Nigeria	C6/36 2	AF425625
Dakar A-1520	IvCo85	1985	Ivory Coast	C6/36 2	AF425620
Rio H 36589	Ango88	1988	Angola	C6/36 4	AF425610
TH-SMAN	Thai54	1954	Thailand	NP	D10513[f]
16007	Thai64	1964	Thailand	Mosq. 2, LLC-MK_2 1	AF180817[f]
West Pac 74	Naur74	1974	Nauru Island	NP	U88535[f]
FGA/89	FrGu89	1989	French Guyana	NP	AF226687[f]
BR-90	Braz90	1990	Brazil	NP	S64849[f]
S275/90	Sing90	1990	Singapore	AP61 3, C6/36 11 to 16	M87512[f]
LAO CH 323	Laos96	1996	Laos	NP	AB003090[f]
D1/H/IMTSSA-ABID/99/1056	IvCo99	1999	Ivory Coast	C6/36 1	AF298807[f]
D1/H/IMTSSA/98/606	Djib98	1998	Djibouti	C6/36 1	AF298808[f]
D1/H/IMTSSA-CAMB/98/658	Camb98	1998	Cambodia	C6/36 1	AF309641[f]
GZ/80	GZ/80	1980	China	?	AF350498[f]
Mochizuki	Japa43	1943	Japan	?	AB074760[f]
A88	Indo88	1988	Indonesia	?	AB074761[f]

[a] Isolation year.

[b] Symbols: C6/36, clone of *Aedes albopictus* cells; SM, suckling mice; Mosq., passage in live mosquitoes (*Toxorhynchites amboinensis* by intrathoracic inoculation); ?, prior passage history unknown; NP, not published.

[c] Accession numbers created by this study except otherwise indicated.

[d] Strains from patients with dengue hemorrhagic fever.

[e] Strains from patients with dengue fever.

[f] Sequences determined in previous studies, retrieved from GenBank.

Deduced amino acid sequences of the E protein of dengue virus 1

The deduced amino acid (aa) sequences of 41 DV-1 E protein sequences were aligned and compared. The E protein sequences of geographically diverse DV-1 strains were highly conserved, with amino acid identities ranging from 95 to 100% (data not shown). Several distinctive features of the E molecule were completely conserved among all DV-1 strains: (a) the 12 cysteine residues involved in disulfide bridge formation; (b) the glycine-rich putative fusion domain of the flavivirus (located

TABLE 2

Mean Nucleotide Distances for the Envelope Gene (E) between Genotypes of Dengue Virus Type 1

Genotype	Distance (SE) (%)[a]	Distance intervals minimum-maximum (%)[b]	No. of strains
I	3.5 (0.3)	1.1–5.8	8
II	2.1 (0.5)	0.6–2.8	3
IV	2.8 (0.3)	0.6–4.0	5
V	2.4 (0.2)	0.2–6.0	26
V (American strains)	1.6 (0.1)	0.2–3.3	22
V (African strains)	3.0 (0.4)	2.2–4.0	3
V (African and Asian strains)[c]	3.4 (0.4)	2.2–4.1	4
Total (all strains)	5.6 (0.3)	0.2–9.4	43

[a] Mean pairwise distance and its standard error for the set of sequences within each genotype. Uncorrected pairwise distances (p-distances) were used.

[b] Minimum distance and maximum distance for the set of sequences within each genotype.

[c] The Singapore 90 strain was excluded from the analysis since considered a recombinant strain (Tolou *et al.,* 2001).

between amino acids 98 and 111), which is universally conserved among DV-2 and DV-3 but not among all DV-4 strains; (c) the predicted N-linked glycosylation sites at positions 67 and 153, with the exception of strain Japa43, which lacked the site at position 67. A total of 74/495 amino acid changes were detected among the strains of

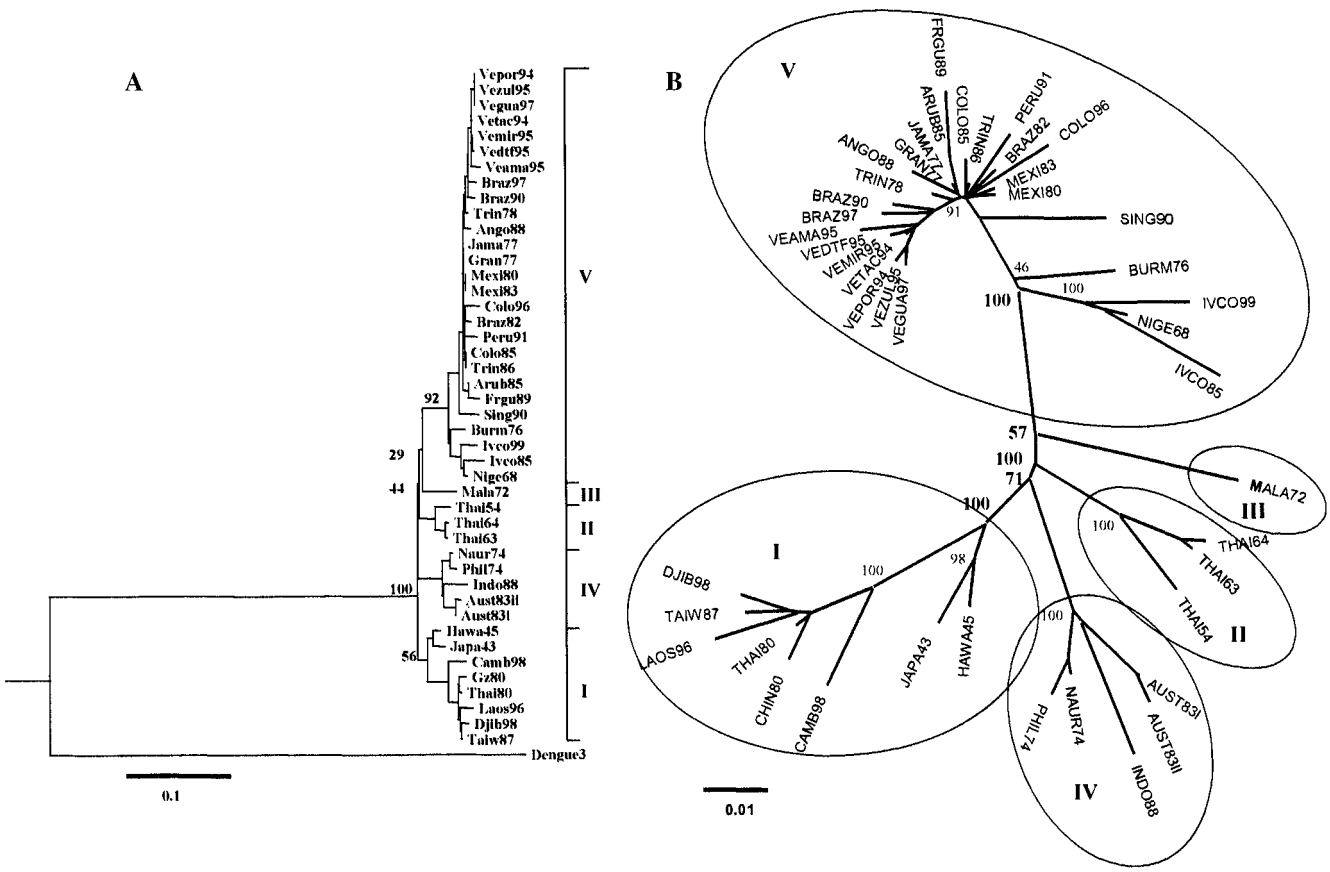

FIG. 1. Maximum likelihood trees showing the phylogenetic relationships among 44 strains of DV-1. The analysis was based on nucleotide sequences of the E gene. The percentage support, based on 100 replicates, is indicated for the principal nodes. In bold, bootstrap values of nodes associated with the definition of genotypes. The scale shows the genetic distance as a percentage of substitutions. (A) The tree is rooted using a representative strain each of dengue 3, dengue 2, and dengue 4 (the last two are not shown in the figure). (B) An unrooted tree is shown. Both trees were constructed with DNAML, using the F84 model of nucleotide substitutions, with a gamma distribution with eight categories, an alpha value of 0.40, and a Ts:Tv ratio of 7.5.

TABLE 3

Calibration of the Molecular Clock for Selected Genetic Groups of DV-1

Type of analysis	Genetic group (No. of strains)	Reference strain (*) or outgroup (**)	dS or dT/tcp[a] ($\times 10^{-4}$)	dN[b] ($\times 10^{-4}$)	dT[c] ($\times 10^{-4}$)	LR test[d]
Regression	American strains (20)	Gren77*	22.2 ± 2.8	0.27 ± 0.6	7.9 ± 1.6	
Regression	American strains (20)	Jama77*	22.3 ± 2.8	0.05 ± 0.6	7.9 ± 1.5	
Regression	Genotype I (7)	Japa43*	24.5 ± 2.0	0.17 ± 0.6	7.8 ± 0.9	
TipDate[e]	All DV-1 strains (43)[f]	Genotype II**			5.8 ± 0.6	$P = 0.0012$
TipDate	All DV-1 strains (43)[f]	Genotype II**	16.2 ± 1.5			**$P = 0.2728$**
TipDate	American strains (22)	Nige64**			7.1 ± 0.9	$P = 0.017$
TipDate	American strains (22)	Nige64**	16.7 ± 2.3			**$P = 0.2423$**
TipDate	Genotype I (9)	Naur74**			4.8 ± 1.0	**$P = 0.1005$**
TipDate	Genotype I (9)	Naur74**	16.7 ± 3.2			**$P = 0.7797$**

[a] Synonymous substitutions/synonymous site/year for the regression results; substitutions/third codon position site/year for the Tip Date results.
[b] Nonsynonymous substitutions/nonsynonymous site/year.
[c] Total substitutions/site/year.
[d] Likelihood ratio test. The *P*-values for which the clock with dated tips model cannot be rejected ($P > 0.05$) are in bold.
[e] All TipDate results shown in the table were obtained using the REV nucleotide substitution model.
[f] Strain Sing91 not included.

DV-1, while only six of these resulted in a side chain difference (polar vs nonpolar) between genotypes and/or groups of DV-1: E-37 (Domain I, Asn → Asp) for genotypes I and IV, E-88 (Domain I, Ala → Thr) for genotype IV, E-161 (Domain I, Thr → Ile) for genotype V, E-297 (between domains I and III, Met → Thr) for some strains from Brazil and Venezuela, E-370 (Domain III, Ala → Thr) for the genotype V strains, and E-478 (*trans*-membrane region, Thr → Met) for genotype II.

Evolution of E gene and the molecular clock

Evolutionary rates for DV-1 were estimated using two different approaches. Assuming that there was one introduction of DV-1 in the Americas in 1977 and that all strains isolated in the following years originated from this putative ancestor, we estimated the substitution rates of the American strains isolated in a 20-year period, compared with the Gren77 strain. Figure 2 shows the relationship between the number of substitutions and the differences in the year of isolation. From the slope of the corresponding line, the synonymous substitutions fixation rate was estimated at $22.2 \pm 2.8 \times 10^{-4}$ synonymous substitutions/synonymous site/year. Instead the number of nonsynonymous substitutions/nonsynonymous site (K_n) was relatively invariable in a 20-year period with an *r* value, indicating no relationship between the variables ($P > 0.76$). Similar results were obtained referring nucleotide substitutions to the Jama77 strain ($K_s = 0.00222$, $r = 0.88$; $K_n = -0.000062$, $r = -0.2532$). An additional estimate of the evolution rate was obtained for strains of genotype I, measuring substitutions of the seven strains to the oldest isolate, Japa43. Similar values of synonymous and total substitution rates were obtained as with the American strains, confirming that different DV-1 genotypes have evolved at similar rates (Table 4). The regression analysis of genetic distances and time, although useful to detect and represent the different behavior of K_s vs K_n, does not allow for a rigorous calibration of the mutation rates, since the points on which the regression analysis was conducted were not completely independent, as they shared some evolutionary history. We then tested whether the DV-1 sequence data would fit a molecular clock model. The single rate with dated tips (SRDT) model was used, as implemented in the TipDate program (Rambaut, 2000), which provides a maximum likelihood estimate of the substitution rate under a molecular clock model, allowing for samples

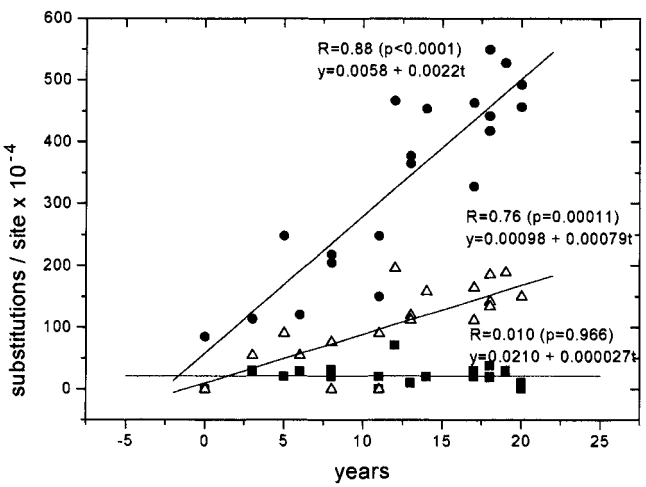

FIG. 2. Relationship between the number of total (△), synonymous (●), and nonsynonymous (■) nucleotide substitutions per site and the difference in the year of isolation for the E gene of the American strains of DV-1. Equations of the regression lines are shown, where "t" is the difference in the year of isolation. Pairwise comparisons of maximum likelihood estimates of each type of substitution rates were made with the Gren77 strain. R, regression constant.

TABLE 4

Primers for Amplification and Sequencing of E Gene of Dengue-1 Virus

Name	Sequence	Position[a]	Use
D1-820-S	5′GAGACACCCAGGATTCACGG3′	820–839	Amplification and sequencing
D1-2600-AS	5′TGGCTGATCGAATTCCACAC3′	2581–2600	Amplification and sequencing
D1-1220-S	5′TTTGTGGACAGAGGCTGGG3′	1220–1238	Sequencing
D1-1559-S	5′CACAAACAATGGTTTCTAGACTTAC3′	1559–1583	Sequencing
D1-1948-S	5′GACCCAAGATGAGAAAGGAGT3′	1948–1968	Sequencing
D1-1240-AS	5′TGCCCCAGCCTCTGTCCAC3′	1222–1240	Sequencing
D1-1424-AS	5′GACGTAGGAGCTTGAGGTGTTAT3′	1402–1424	Sequencing
D1-1868-AS	5′CATGCTGGGTCTCAGCCAC3′	1850–1868	Sequencing

[a] Based on the numbering of Nauru strain (Accession No. U88535).

isolated at different dates. The 43 DV-1 strains were used after excluding the putative recombinant strain Sing91 (Tolou *et al.,* 2001). Maximum likelihood phylogenies were obtained under a standard model (different rate, DR) and the dated tip model. Conditions tested included different models of substitutions (F84, HKY85, REV) and rooting with different outgroups. The likelihood ratio test did not support a molecular clock for the five genotypes on the basis of the total nucleotide substitution rate but supported it when substitution rates were based on the third codon position (Table 3). Analysis by the same approach of different subsets of strains (genotypes I and V) confirmed that substitution rate values fell within a close range. A tree constructed under the assumption of a molecular clock with dated tips, using the estimate of $16.2 \pm 1.5 \times 10^{-4}$ substitutions/third position codon site/year, is shown in Fig. 3. The root of the tree, corresponding to the time of divergence of genotypes I, II, and IV, fell at the year 1906 (95% confidence interval: 1900–1912). Then, divergence of four genotypes of DV-1 (I–IV) and of the putative ancestor of genotype V, was calculated to occur at the beginning of 1900 (1900–1916), whereas divergence of genotype V was estimated to take place approximately 40 years later (1945–1953). The ancestral node for the American strains, first isolated in the continent in 1977, was calculated to have occurred at 1973 (±2 years), confirming, for a well-documented epidemiological situation, the realism of the predictions. On the whole these values corresponded very closely to those calculated using the dS values obtained by regression analysis (Table 3) and a synonymous substitution tree: with these parameters, divergence of genotypes I to IV was estimated to have occurred between 1898 and 1934, whereas genotype V had diverged between 1945 and 1960 (data not shown).

Genetic polymorphism of the E gene of dengue virus 1

The number of synonymous vs nonsynonymous substitutions per site for different regions of the E gene was estimated for a subset of 23 strains representative of the five genotypes. Eight regions were individually analyzed. These corresponded to the transmembrane domain plus the three ectodomains, each subdivided into the corresponding discontinuous segments, assuming for DV-1 a structure similar to the one of the soluble portion of the E protein of tick-borne encephalitis virus (Rey *et al.,* 1995). In the various regions the values of synonymous substitutions rates (K_s) ranged between 13 and 37 times the values of nonsynonymous substitution rates (K_n) (data not shown). The excess of synonymous substitutions in all the regions of the molecule indicates a strong purifying selection on E gene of DV-1.

DISCUSSION

Genetic diversity of DV-1 strains has been analyzed in previous studies. T1 fingerprint analysis allowed their first classification in eight topotypes (Trent *et al.,* 1990). Rico-Hesse (1990) sequenced a fragment of 240 bases in the E/NS1 junction of 40 DV-1 strains and classified them in five genotypes: the first group contained virus from the Americas, Africa, and Southeast Asia; the second included a single isolate from Sri Lanka; the third a strain from Japan isolated in 1943; the fourth included strains from Southeast Asia, the South Pacific, Australia, and Mexico; and the fifth group contained viruses from Taiwan and Thailand. The E/NS1 junction sequences represented approximately 2% of the complete genome of DV-1 and showed around a 9% divergence. Chungue *et al.* (1995) compared a 180-base-pair fragment of the E gene from 35 DV-1 strains; they observed a 7% divergence and classified these strains into three genotypes. Since the two studies included mostly different strains, it is difficult to exactly establish the equivalence between the two classifications. In our study, the phylogenetic analysis of the E gene of 44 DV-1 isolates produced evidence of the existence of five different genotypes. The three genotypes indicated as I, IV, and V are equivalent to the ones referred by Rico-Hesse (1990), with variations in the classification of the Mexi80 and Hawa45 strains, included by Rico-Hesse with strains from Australia and Nauru Island into a genotype corresponding to our ge-

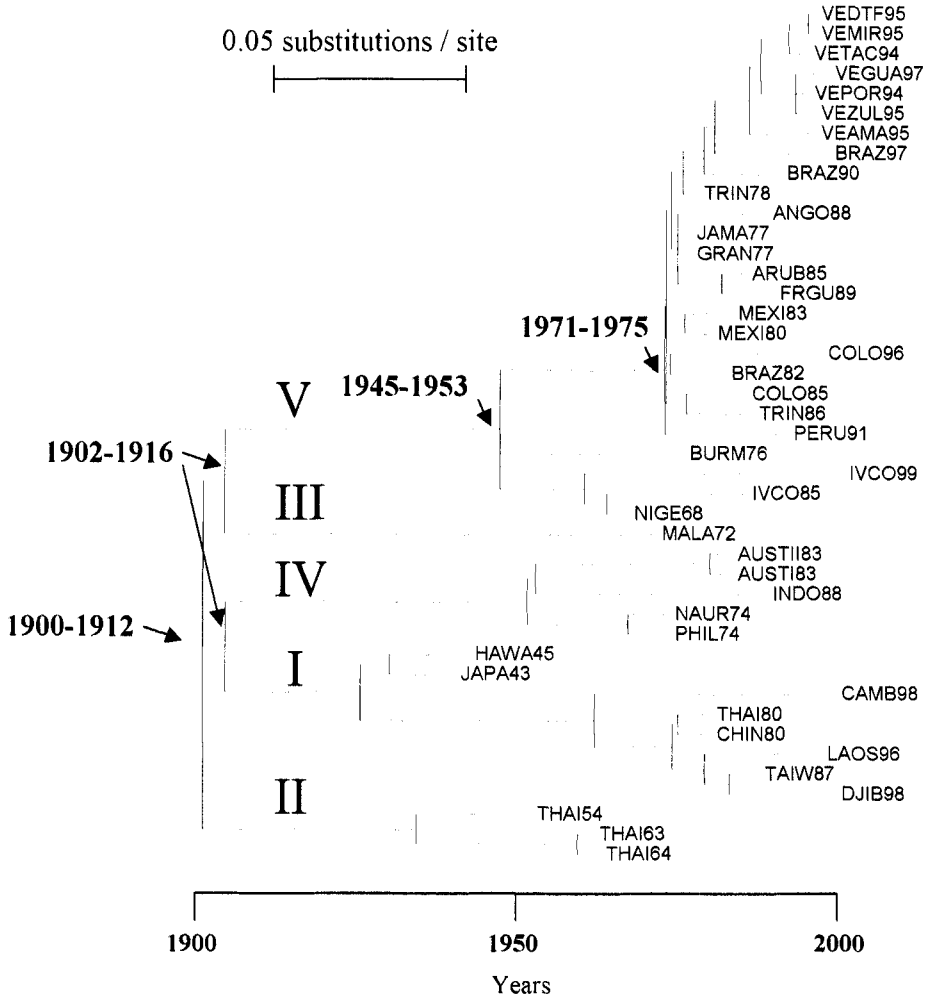

FIG. 3. Time of divergence of different genotypes of DV-1. Estimated divergence times of the nodes originating the genotypes of DV-1. Dates, expressed as the range corresponding to a 95% confidence interval, are indicated by arrows on a maximum likelihood tree, constructed under the single rate-dated tips model (TipDate; Rambaut, 2000). The absolute time scale is aligned to the root of the tree, which corresponds to the year 1906. The branch lengths are expressed as substitution/third codon position site.

notype IV. In our analysis this Mexican strain exhibited a high identity with the other American isolates (genotype V), whereas the Hawa45 strain was included in genotype I with strains from Southeast Asia. The nucleotide sequences of E gene of the strains Mexi83 (1378) and Arub85 (495-1) obtained by us were different from those reported for the same strains (D005004 and D00505, respectively) by Chu *et al.* (1989). Interestingly pairs of sequences for each strain were virtually identical (0.6% divergence) in the region comprising nucleotides 1–312 of the E gene, but strongly diverged (6.4–6.6%) from nucleotides 313–1485 (data not shown). The sequences published by these authors of the E gene of the American isolates display a high identity with strains from Thailand, Philippines, and with Nauru, so that they would be classified as genotype IV in the scheme proposed in Fig. 1. This would indicate the circulation of two different genotypes in the Americas. In our interpretation the sequences of the three American strains (Mexico83, Aruba85, and Jamaica77) reported by Chu *et al.* (1989) are not reliable. If our interpretation is correct, the occurrence of recombination among DV-1 strains, which has been based on these sequences (Holmes *et al.,* 1999; Worobey *et al.,* 1999), must be reconsidered. The present results indicate that from the first introduction in 1977 only one genotype of DV-1, with the potential to cause FD and FHD, has circulated in the Americas.

Although prospective studies suggest that the risk for DHF/DSS is greater in infections with DV type 2 (Burke *et al.,* 1988), it has been reported in several studies that DHF/DSS can occur during infection with any of the four serotypes of DV (reviewed by Rothman, 1997). There are no reports that associate particular strains and/or genotypes of serotype DV-1 with disease severity. The American genotype, the most extensively studied by us, included Venezuelan strains able to cause DF or DHF, which did not segregate with the clinical disease severity. This finding suggests that

this genotype has the potential to cause the most severe form of the disease.

The evolution of the E gene of DV-1 was shown to comply with a molecular clock when substitutions were measured at the third codon position. This type of substitution is contributed to mainly by synonymous substitutions and as such is less influenced by a positive or a purifying selection pressure. Following a molecular clock model implies that the different genetic lineages of DV-1 had evolved at a constant rate despite the fact that lineages were associated to a world-wide geographical spread, with different epidemiological situations involved, such as bottlenecks and sustained periods of epidemic or endemic transmissions. The evolution rate of DV-1 expressed as total substitutions (5.8×10^{-4}/site/year) is very similar to the one reported for DV-4, by Lanciotti *et al.* (1997) ($8–10.4 \times 10^{-4}$ substitutions/site/year, estimated by regression analysis) and, on the same data set, by Rambaut (2000) (7.9 ± 1.9 substitutions/site/year estimated according to the SRDT model). Thus, at least for these two serotypes, dengue virus appears to have evolved at a similar rate. This might not be so for the other serotypes. For instance a rather lower rate (5×10^{-4} synonymous substitutions/synonymous site/year) has been estimated for DV-2, based on the analysis of a clade of viruses from South-East Asia (Wang *et al.,* 2000).

Based on the estimated evolution rate, the cladogenesis of the DV-1 genotypes I to IV considered in this study was estimated to have occurred approximately 100 years ago. Epidemics attributed to dengue viruses have occurred at least in the last three centuries. In the specific case of DV-1, its circulation in several parts of the world in the first half of the 20th century has been demonstrated by serological evidence (Gubler, 1997). Possibly a virus strain associated with a world pandemic constituted the single ancestral lineage from which genotypes I to IV of DV-1 originated. These genotypes have been then circulating in Asia, Oceania, and Africa in the last 50 years.

At the protein level, no accumulation of nonsynonymous substitutions and therefore no cumulative fixation of amino acid changes occurred throughout a 20-year period, as shown by the evolution of the American strains. Virus genes generally behave as predicted by the neutral theory, the protein sequence evolution being under purifying selection and random genetic drift, variation being limited by structural and functional constraints (Kimura, 1983; Gojobori *et al.,* 1990). The evolution of E gene of DV-1 is consistent with this model, as indicated by the complete absence of accumulation with time of fixed mutations at the protein level, indicative of a strong purifying selection. This was confirmed by the analysis of different regions of the E gene, corresponding to the different structural and functional domains of the molecule. In all regions, the nonsynonymous/synonymous substitution rates ratio ($\omega = K_n/K_s$) had $\omega \ll 1$ values, indicating a predominant purifying selection. These findings are in agreement with those reported by Yang *et al.* (2000), who also demonstrated very low ω values (approximately 0.06) for the E gene of selected DV of the four serotypes. These authors, applying an analysis based on the measurement of ω for all codons of a gene over a phylogeny, did not find evidence for positive selection in the E gene of neither DV nor Japanese encephalitis virus (Yang *et al.,* 2000). We applied additional tests to ascertain a positive selection contribution to the polymorphism in E gene of DV-1. Results of the Tajima test and the Fu and Li test (Tajima, 1989; Fu and Li, 1993) indicated a nonsignificant departure from neutral expectations for the E gene ($P > 0.10$) (data not shown). Thus a role for positive selection in maintaining the diversity of the E gene, favoring for instance phenotypes with an increased capability to escape the immune response, as proposed for other systems (Ina and Gojobori, 1994; Yang *et al.,* 2000; Haydon *et al.,* 2001) or with modified patterns of transmission (Zanotto *et al.,* 1996; Holmes *et al.,* 1998), remains to be established.

MATERIALS AND METHODS

Virus

DV-1 strains sequenced in this article included the following: (a) seven isolates from Venezuela, consisting of the first passage in C6/36 cells of the original patient serum; and (b) a collection of 24 DV-1 strains representative of different regions of the world and years of isolation, kindly provided by Dr. Robert Tesh (University of Texas, Galveston) (Table 1). These isolates were used as the next C6/36 mosquito cell passage of the virus seeds with the history described in Table 1. All these viruses were isolated from human sera, with the exception of strain P72-1244 (D1-Mal-72) isolated from monkeys.

RNA extraction, RT, and PCR

The entire E gene of DV-1 (1485 nucleotides) was amplified using RT-PCR. RNA was extracted from 200 μl of cell-culture supernatant using Trizol LS (Invitrogen), following the manufacturer's instructions. Dry RNA pellets were resuspended in a mix of water:DMSO (3.6:1), heated at 95°C for 5 min, and cooled on ice. The sample (10 μl) was added to the RT mix: buffer Superscript II RT (50 mM Tris–HCl, 75 mM KCl, 3 mM $MgCl_2$, pH 8.3; Invitrogen), 10 mM DTT, 333 μM of the antisense primer (Table 4), 0.66 mM of each dideoxynucleotide (Promega), 40 U Rnasin (Promega), and 200 U of reverse transcriptase (Superscript II, Invitrogen). The mixture was incubated at 42°C for 1 h and then at 50°C for 1 h 30 min. After heating at 70°C for 15 min, the mixture was treated with ribonuclease H at 37°C for 20 min. The mixtures were kept at −20°C until use. For PCR amplification, 2 μl

cDNA was added to a mixture containing 400 μM of the sense and antisense primers for PCR (Table 4), PCR buffer (10 mM Tris–HCl, 50 mM KCl, 0.1% Triton X-100), $MgCl_2$ 2 mM, 75 μM of each deoxinucleotide and 2.5 U Taq DNA polimerase (Promega). Samples were amplified for 30 cycles at 94°C for 45 s, 55°C for 45 s, and 72°C for 2 min.

Sequencing of PCR fragments

For automated sequencing, spin column-purified (Concert Rapid PCR Purification System, Invitrogen) DNA fragments were analyzed by the cycle-sequencing dye terminator method (The Big Dye Terminator Cycle Sequencing Ready Reaction kit, Perkin–Elmer, Applied Biosystems, Foster City, CA). For each sequencing reaction, approximately 50 to 100 ng purified DNA was combined with 3.7 pmol of sequencing primer (Table 4) and a reaction cocktail containing the four dye-labeled dideoxynucleotide terminators. Cycle-sequencing parameters used were as described in the manufacturer's protocol. Samples were analyzed in an Applied Biosystems Prism 377 sequencer (Applied Biosystems, Perkin-Elmer). Overlapping nucleic acid sequences were combined for analysis and edited with the software Autoassembler (ABI PRISM, Applied Biosystems, Perkin-Elmer).

Phylogenetic analysis

The multiple sequence alignment program Clustal W (version 1.6) (Thompson *et al.,* 1994) was used to obtain an optimal nucleotide sequence alignment file. Phylogenetic trees for the entire sequence of E gene were obtained using different tree building methods either by MEGA (version 2.1) (Kumar *et al.,* 2001) or PHYLIP (version 3.6 alpha 2) (Felsenstein, 1993). Various genetic distance matrices were used with the neighbor-joining method (NJ) (Saitou and Nei, 1987), calculating bootstrap confidence intervals of 1000 replicates. Character state tree-building algorithms (PHYLIP package) were also tested. A strict consensus bootstrap tree was obtained by using the following programs: (i) SEQBOOT to generate 100 replicas, (ii) DNAML or DNAPARS to acquire the tree of each reiterated data, and (iii) CONSENSE to build a strict consensus bootstrap tree. The DNAML model of DNA substitution (F84) was used with rate variation among sites, approximating a Gamma distribution, with eight rate categories. The shape parameter (alpha) of the distribution and the transition:transversion ratio were calculated using the Baseml program from the PAML 3.12 package (Yang, 1997). Phylogenetic trees were drawn using Treview (Page, 1996) or TreeExplorer (http://evolgen.biol.metro-u.ac.jp/TE/TE_man.html). Trees were rooted using the homologous sequences of either one dengue serotype 3 strain (L11428) or three strain representative of serotypes 3 (L11428), 2 (AF100468), and 4 (U18425).

Estimation of dengue 1 divergence rates

The time of divergence of DV-1 lineages were estimated using the following two different approaches.

(a) regression analysis, with the two more numerous DV-1 genotypes. Pairwise synonymous, nonsynonymous, and total substitution per respective site were calculated (i) for a group of 20 sequences corresponding to the American DV-1 strains isolated in a 20-year period in relation to the sequences of the oldest isolates of the group, strains Gren77 or Jama77, or (ii) for seven sequences of genotypes I compared to the oldest isolate Japa43. The substitution values for each sequence were plotted against the difference in the year of isolation. Regression curves were constructed with the Micrococal Origin software (Northampton, MD), with correlation coefficients (r values) calculated by the least-squares method. For these calculations maximum likelihood estimates of the synonymous substitutions per synonymous site (K_s) and nonsynonymous substitutions per nonsynonymous site (K_n) among pairs of sequences were used, obtained by the method of Goldman and Yang (1994), as implemented in the Codeml program of the PAML package, version 3.12 (Yang, 1997). Divergence times for specific nodes were calculated on a synonymous substitution distance tree, constructed with the Neighbor program of Phylips on the pairwise synonymous substitution matrix. Distances from the node of interest to each tip of the clade were divided by the synonymous substitution rate and the resulting time value were subtracted from the year of isolation of each strain.

(b) Rates of nucleotide substitutions per site or per codon position were also estimated using the TipDate program (Rambaut, 2000), implemented as one of the molecular clock options in Baseml program of PAML or accessed at http//bioweb.pasteur.fr/seqanal/tmp/tipdate. TipDate provides a maximum likelihood estimate of the evolution rate of sequences isolated at different dates, assuming a molecular clock (constant rates of substitutions) but allowing for different dates of isolation. The program estimates also the dates and the corresponding confidence intervals of the ancestor nodes of the phylogeny. For each chosen set of sequences to be analyzed, a maximum likelihood tree was first obtained with DNAML (Phylips) and then used as input tree with the corresponding alignment in Baseml. Each sequence set was tested under different nucleotide substitution models and with two different clock options: either the nonclock (different rate) or the clock with dated tips model (SRDT). Analysis was performed for either total substitutions or for substitutions at each codon position. The validity of the SRDT model against the nonclock DR model for each data set was estimated using the likelihood ratio test (Felsenstein, 1993). All results referred to were obtained with trees constructed with the REV+Γ substitution model (Yang, 1997) with eight rate categories and with the alpha and kappa parameters estimated from the data.

ACKNOWLEDGMENTS

We thank Dr. Robert Tesh for generously providing the DV-1 strains. This work was supported by a grant from Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICIT; G-98002081) to F.L. and R.S. A.P.G. was supported by a doctoral fellowship of the University of Carabobo. We thank Pulcherie Guenau, Roberto Cipriani, and Omar Cornejo for generous help and suggestions with the data analysis. The personnel of CeSAAN is gratefully acknowledged for assistance with sequencing. CeCalcULA is thanked for assistance with computing procedures.

REFERENCES

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

Chu, M. C., O'Rourke, E. J., and Trent, D. W. (1989). Genetic relatedness among structural protein genes of dengue 1 virus strains. *J. Gen. Virol.* **70,** 1701–1712.

Chungue, E., Cassar, O., Drouet, M. T., Guzman, M. G., Laille, M., Rosen, L., and Deubel, V. (1995). Molecular epidemiology of dengue-1 and dengue-4 viruses. *J. Gen. Virol.* **76,** 1877–1884.

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) version 3.6a2. Distributed by author. Department of Genetics, University of Washington, Seattle, WA.

Fu, Y. X., and Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* **133,** 693–709.

Gojobori, T., Moriyama, E. N., and Kimura, M. (1990). Molecular clock of viral evolution, and the neutral theory. *Proc. Natl. Acad. Sci. USA* **87,** 10015–10018.

Goldman, N., and Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11,** 725–736.

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" (D. J. Gubler and G. Kuno, Eds.), pp. 1–22. CAB International, New York.

Halstead, S. B. (1988). Pathogenesis of dengue: Challenges to molecular biology. *Science* **239,** 476–481.

Haydon, D. T., Bastos, A. D., Knowles, N. J., and Samuel, A. R. (2001). Evidence for positive selection in foot-and-mouth disease virus capsid genes from field isolates. *Genetics* **157,** 7–15.

Holmes, E. C., Bartley, L., and Garnett, G. P. (1998). The emergence of dengue: Past, present, and future. *In* "Emerging Infections." pp. 301–304. Academic Press, San Diego.

Holmes, E. C., Worobey, M., and Rambaut, A. (1999). Phylogenetic evidence for recombination in dengue virus. *Mol. Biol. Evol.* **16,** 405–409.

Ina, Y., and Gojobori, T. (1994). Statistical analysis of nucleotide sequences of the hemagglutinin gene of human influenza A viruses. *Proc. Natl. Acad. Sci. USA* **91,** 8388–8392.

Kimura, M. (1983). "The Neutral Theory of Molecular Evolution." Cambridge Univ. Press, Cambridge, U.K.

Kumar, S., Tamura, K., Jacobsen, I., and Nei, M. (2001). "Molecular Evolutionary Genetics Analysis Software." Arizona State Univ., Tempe, AZ.

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

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

Leitmeyer, K. C., Vaughn, D. W., Watts, D. M., Salas, R., Villalobos de Ch., I., Ramos, C., and 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., and Trent, D. W. (1993). Phylogenetic relationships of dengue-2 viruses. *Virology* **197,** 216–224.

Monath, T. P. (1994). Dengue: The risk to developed and developing countries. *Proc. Natl. Acad. Sci. USA* **91,** 2395–2400.

Monath, T. P., and Heinz, F. X. (1996). Flaviviruses. pp. 961–1034. *In* "Fields Virology" (B. N. Fields, D. M. Knipe, and P. M. Howley, Eds.), Lippincott-Raven, Philadelphia, New York.

Page, R. D. M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput. App. Biosci.* **12,** 357–358.

Pinheiro, F. P. (1989). Dengue in the Americas. 1980–1987. *Epidemiol. Bull.* **10,** 1–8.

Pinheiro, F. P., and Corber, S. J. (1997). Global situation of dengue and dengue haemorrhagic fever and its emergence in the Americas. *World Health Stat. Q.* **50,** 161–169.

Rambaut, A. (2000). Estimating the rate of molecular evolution: Incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* **16,** 395–399.

Rey, F. A., Heinz, F. X., Mandl, C., Kunz, C., and Harrison, S. C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* **375,** 291–298.

Rico-Hesse, R. (1990). Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. *Virology* **174,** 479–493.

Rico-Hesse, R., Harrison, L. M., Salas, R. A., Tovar, D., Nisalak, A., Ramos, C., Boshell, J., de Mesa, M. T., Nogueira, R. M., and da Rosa, A. T. (1997). Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. *Virology* **230,** 244–251.

Rosen, L. (1977). The Emperor's New Clothes revisited, or reflections on the pathogenesis of dengue hemorrhagic fever. *Am. J. Trop. Med. Hyg.* **26,** 337–343.

Rothman, A. L. (1997). Viral pathogenesis of dengue infections. *In* "Dengue and Dengue Hemorrhagic Fever" (D.J. Gubler and G. Kuno, Eds.), pp. 245–271. CAB International, New York.

Salas, R. A., Tovar, D., Barreto, A., de Miller, E., Leitmeyer, K., and Rico-Hesse, R. (1998). Serotipos y genotipos de virus dengue circulantes en Venezuela, 1990–1997. *Acta Cient. Venezol.* **49,** 33–37.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **6,** 514–525.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123,** 585–595.

Thompson, J. D., Higgins, D. G., and 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 Res.* **22,** 4673–4680.

Tolou, H. J. G., Couissinier-Paris, P., Durand, J. P., Mercier, P., de Pina J. J., de Micco, P., Billoir, F., Charrel, R. N., and de Lamballerie, X. (2001). Evidence for recombination in natural populations of dengue virus type 1 based on the analysis of complete genome sequences. *J. Gen. Virol.* **82,** 1283–1290.

Trent, D. W., Manske, C. L., Fox, G. E., Chu, M. C., Kliks, S. C., and Monath, T. P. (1990). The molecular epidemiology of dengue viruses: Genetic variation and microevolution. *Appl. Virol. Res.* **2,** 293–315.

Wang, E., Ni, H., Xu, R., Barret, A. D. T., Watowich, S. J., Gubler, D. J., and 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., and Halstead, S. B. (1999). Failure of secondary infection with American genotype dengue 2 to cause dengue hemorrhagic fever. *Lancet* **354,** 1431–1434.

Worobey, M., Rambaut, A., and Holmes, E. C. (1999). Widespread intraserotype recombination in natural populations of dengue virus. *Proc. Natl. Acad. Sci. USA* **96,** 7352–7357.

Yang, Z. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13,** 555–556.

Yang, Z., Nielsen, R., Goldman, N., and Krabbe, P. A.-M. (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., and Holmes, E. C. (1996). Populations dynamics of flaviviruses revealed by molecular phylogenies. *Proc. Natl. Acad. Sci. USA* **93,** 548–553.