

## **NS 1 gene sequences from eight dengue-2 viruses and their evolutionary relationships with other dengue-2 viruses**

**J. Blok<sup>1</sup>, A. J. Gibbs<sup>2</sup>, S. M. McWilliam<sup>1, \*</sup>, and U. T. Vitarana<sup>3</sup>**

<sup>1</sup>Sir Albert Sakzewski Virus Research Laboratory, Royal Children's Hospital,  
Brisbane, Queensland

<sup>2</sup>Research School of Biological Sciences, Australian National University, Canberra, ACT,  
Australia

<sup>3</sup>Medical Research Institute, Colombo, Sri Lanka

Accepted November 11, 1990

**Summary.** The nucleotide sequences of the NS 1 genes from five Thai and three Sri Lankan dengue-2 viruses were determined by sequencing the viral RNA using synthetic oligonucleotide primers. The results were shown to be similar to four published dengue-2 NS 1 sequences and the classification of these genes was compared with the one obtained for the envelope genes of the same viruses. The classification was similar and showed that the Thai isolates could be divided into two separate groups and that the Sri Lankan isolates were distinct. We found no correlation between disease severity, serological response (1° or 2°), or year of isolation and various aspects of NS 1 protein sequence variation; and no particular amino acid changes were correlated with virulence. The sequences were combined with those published and classified elsewhere to provide a comprehensive E/NS 1 gene taxonomy of dengue-2 virus isolates.

### **Introduction**

Classical dengue fever, characterized by fever, rash, headache and arthralgia occurs in many parts of the world including South-East Asia, the western Pacific islands, the Caribbean and East Africa. The more severe form of the disease, dengue haemorrhagic fever (DHF) is characterized by haemoconcentration and a haemorrhagic diathesis which can lead to dengue shock syndrome (DDS) and may be fatal. This severe form of the disease occurs in various places including South-East Asia, the Caribbean and western Pacific. The number of DHF/DSS cases in different areas is, however, not the same. For instance, in Thailand

---

\* Present address: CSIRO Division of Tropical Animal Production, Long Pocket Laboratories, Indooroopilly, Qld, Australia.

over 20,000 patients with DHF are hospitalized annually [31] while in Sri Lanka, about 10 DHF cases have been recorded annually [5]. However, in 1989 and the first half of 1990, the number of DHF cases in Sri Lanka has increased to 290 confirmed cases of which 61 were fatal, even though the mosquito vector densities or the number of dengue fever cases have remained similar to previous years.

Dengue is caused by a flavivirus, a family of viruses with RNA genomes of about 11,000 nucleotides [25] which are transmitted by ticks and mosquitoes [37]. There are four dengue virus serotypes, all of which have been associated with the mild and severe forms of the disease [33]. All four serotypes are present in both Thailand and Sri Lanka, and the incidence of sequential infections appears to be similar. The reason for the difference in the number of DHF cases in these two countries is not clearly understood but it may be due to differences in virus virulence.

Several studies using RNase T<sub>1</sub> oligonucleotide fingerprinting [18, 24, 34–36], restriction enzyme mapping [4, 36] or antigenic signature analysis using monoclonal antibodies [22] have shown that dengue viruses within one serotype can vary. This variation has allowed some geographic groupings to be determined but no correlation with disease severity has been found.

A limited study of the gene sequence coding for the major antigenic protein of dengue viruses, the envelope protein found no correlation between sequence variation and disease severity [5]. Another study using a 240 nucleotide long region of the envelope and NS 1 gene junction [26] was unable to obtain a correlation between disease severity and features of the sequences, due to the lack of isolates representing the various disease severities. However, both studies found differences between isolates and dendrograms to illustrate the relationships of the strains were calculated.

It is possible that the epidemiology and pathogenic potential of dengue viruses is determined as much by the non-virion proteins as by those which form part of the virion. Therefore, gene sequences encoding the NS 1 protein were studied. The function of the NS 1 protein is unknown but it was originally detected as a soluble complement fixing antigen in the media around cultured cells infected with dengue virus [7] and in the serum of infected mice [6]; and it has also been identified on the surface of infected cells [8]. The role of NS 1 in viral replication is unknown but it has been suggested that it may function in virus assembly and maturation [25].

Immunization with NS 1 protein or parenteral injections of anti-NS 1 monoclonal antibodies have been found to protect mice from lethal challenge with dengue-2 virus [14, 32] suggesting that this protein may play an important role in dengue infections. Studies with *in vitro* expressed NS 1 [23] have allowed the determination of functional domains as well as antigenic domains at the N- and C-termini. Monoclonal antibodies to N- and C-terminal peptides did not however protect mice [23].

We report in this paper the NS1 gene sequences from the same five Thai and three Sri Lankan dengue-2 virus isolates used in the envelope gene study [5], and compare these with other published dengue-2 NS1 sequences from viruses isolated from the Caribbean [9, 12] as well as the prototype New Guinea C strain [16].

### Materials and methods

#### *Virus isolates*

The full names, codes, serology, and clinical symptoms caused by the dengue-2 viruses used in this study are listed in Table 1. The 16681 strain was isolated from a fatal case of DSS in Bangkok in 1964 [13], whereas the other Thai viruses (D, F, G, H) were isolated from blood specimens from patients at Bangkok Children's Hospital in 1980 [36] and these were kindly provided by Dr. Don Burke. Clinical diagnosis and grades of disease severity were assigned according to WHO criteria [38] by Dr. Suchitra Nimmannitya. The three Sri Lankan virus isolates were obtained from patients with dengue fever symptoms in 1968, 1969, and 1982. Two of the strains (b, c) were isolated from patient's serum by passage in suckling mice (two to four passages) followed by two passages in mosquito cells; while the other strain (a) was obtained by mosquito (*Aedes aegypti*) inoculation followed by two passages in mosquito cells.

None of the virus stocks were plaque purified before growth in the mosquito cell line C6/36, and the number of passages in this cell line was minimized (less than four passages). In this way we tried to ensure that the virus population present in the original patient dominated the viral genomes we sequenced.

#### *Growth and purification of viruses and genomic RNA*

Viruses were propagated in a continuous line of *Aedes albopictus* cells (clone C6/36) [15]. The cells were grown at 32 °C in RPMI 1640 medium containing 10% foetal calf serum, 100 units/ml penicillin and 100 µg streptomycin, but were placed in medium containing

**Table 1.** Characteristics and codes of the dengue-2 viruses used in this study

Virus isolate	Code	Year of isolation	Country of origin	Clinical disease	Patient age	Sero response	NS1 sequence references
New Guinea C	NGC	1944	New Guinea	DF	adult	—	[16]
16681	16681	1964	Thailand	DSS	8	2°	*
D80-030	D	1980	Thailand	DHF	3	2°	*
PUO-280	F	1980	Thailand	DF	7	2°	*
D80-100	G	1980	Thailand	DHF	4/12	1°	*
D80-141	H	1980	Thailand	DSS	7	2°	*
PR159(S1)	S1	1969	Puerto Rico	Vaccine	—	—	[12]
D2-1409	JAM	1983	Jamaica	—	—	—	[9]
SL767	a	1982	Sri Lanka	DF	11	2°	*
SL77/69	b	1968	Sri Lanka	DF	21	2°	*
SL1050	c	1969	Sri Lanka	DF	50	1°	*

DF Dengue fever, DHF dengue haemorrhagic fever, DSS dengue shock syndrome

\* This paper

16681 GAUAGUGGUUGCGUUGUGAGCUGGAAAAACAAGAACUGAAAUGGGCAGUGGGAUUUUAUCACAGACAACGUGCACACAUUGGACAGAA<sup>90</sup>  
D C C U  
F A G U  
G U U U  
H U G G U U U U  
a U G U U U  
b G U U  
c UG U U  
UG U

16681 CAAUACAAGUCCAACCAGAAUCCCCUCAAACUAGCUUCAGCUAUCCAGAAAGCCCAUGAAGAGGGCAUUUGUGAAUCCGCUCAGUA<sup>180</sup>  
D A U A G  
F A U U A G  
G C U U A G  
H U U A G G U  
a U U G G G  
b U U U U G  
c U U U U U G

16681 ACAAGACUGGAGAUCUGAUGGAAACAAAUAACACCAGAAUUGAAUCACAUUCUACAGAAAUGAGGUGAAGUUAACUUAUUGACA<sup>270</sup>  
D U C G A G C  
F U U G G A G C C  
G U U G G A G C C  
H C A G G C C C  
a C A G G C C C  
b C A G G C C C  
c C G C C C

16681 GGAGACAUAAGAAUACUGCAGGCAGGAAAAACGAUCUCUGCGGCCACAGCCACUGAGCUGAAGUUAUCAUGGAAAAACUUGGGCAA<sup>360</sup>  
D U U U C G G U A G  
F U U U C G G G U A G  
G U U A U C C U C A G  
H U U U U U A G C U C A G  
a U U U U U U G C U C G A G  
b U U U U U U C U G U G  
c U A A G C U G

16681 GCAAAAUGCUCUCUACAGAGUCUCAUAACCAGACCUUUCUCAUUGAUGGCCCGAAACAGCAGAAUGCCCCAACAAAAUAGAGCUUGG<sup>450</sup>  
D G C C C A U C  
F G C C C A U  
G G C C C A U  
H G C C C A U  
a G C C C A U  
b G C C C A U  
c G A G C C C

16681 AAUUCGUUGAAGUUAAGACUAUGGCUUUGGAGUAUUCACCACCAUAUAUUGGCUAAAUGAAGAAAAACAGAUUAUUCUGCGAC<sup>540</sup>  
D C C C U U G G G G U U  
F C C C U U G G G G U U  
G C C C U U G G G G U U  
H C C C U U G G G G U U  
a C C C U U G G G G U U  
b C C C U U G G G G U U  
c C C C U U G G G G U U

16681 UCAAAAUCUAGUCAGCGGCCAUAAGACAAACAGAGCCGUCAUUGCCGAUAUUGGUUAUUGGUAAGAAAGUGCACUAUUGACACAUGG<sup>630</sup>  
D U U C G A U U U U  
F U U C G A U U U U  
G U U C G A U U U U  
H U U C G A U U U U  
a U U C G A U U U U  
b U U C G A U U U U  
c U U C G A U U U U

16681 AAGAUAGAGAAAGCCUUCUUAUGAAGUCAAAAAUCGCCACUGGCCAAAUCACACACCUCUGGACAAUGGAGUCUAGAAAGUGAG<sup>720</sup>  
D G G U U G G  
F U U U U G G  
G U U U U G G  
H A G U U G G A G  
a A U C U G G U G  
b A A U G G U G  
c G A A U G G U G

16681 AUGAUAAUUCCAAAGAUCUCGUGGACCAUGUCUCAACACAACUAUAGACCAGGCUACCAUACACAAAUAACAGGACCAUGGCAUCUA<sup>810</sup>  
D U A U U G A A C C U U  
F U U U U C C U U  
G A U U U G A A C C U U  
H A U U G A A C C U U  
a U U U G A A C C U U  
b U U U G A A C C U U  
c U U U G A A C C U U

16681 GGUAGCUUGAGAUGGACUUGAUUUCUGUGAUGGAACAACAGUGGUAGUGACUGAGGACUGCGGAAAUGAGAGGCCUUCUUGAGAACA<sup>900</sup>  
D C A C C A U A G  
F C A C C A U A G  
G C A C C A U A G  
H C A C C A U A G  
a C A C C A U A G  
b C A C C A U A G  
c C A C C A U A G

16681 ACCACUGCCUCUGGAAAACUCAUAACAGAAUUGGUGCUGCCGAUCUUGCAUAUACCACCGCUAAGAUACAGAGGUGAGGAUGGGUGCUGG<sup>990</sup>  
D U U C A A C  
F U U C A A C  
G U U C A A C  
H U U C A A C  
a U U C A A C  
b U U C A A C  
c U U C A A C

16681 UACGGGAUGGAAUACAGACCAUUGAAGGAGAAAGAGAAUUUGGUCUACUCCUUGGUCACAGCU<sup>1056</sup>  
D A A U A C C  
F G A A U A C C  
G A A U A C C  
H A A U A C C  
a A A U A C C  
b A A U A C C  
c A A U A C C

0.2% bovine serum albumin after the dengue viruses were inoculated. Virus inoculation and daily haemagglutination monitoring were as described [36], and after viral haemagglutinin had been detected, the tissue culture medium containing the mature virions was collected for three consecutive days and centrifuged to remove any cells and debris. Virus was harvested from this medium by precipitating with polyethylene glycol (MW 6000, 6.6% w/v) and then further purified by centrifuging in 5–50% sucrose gradients. RNA was extracted from the viral pellets as previously described [3], and was ethanol precipitated twice before use in sequence experiments.

#### *Nucleotide sequencing*

RNA from the dengue-2 viruses was sequenced using the dideoxynucleotide chain termination method [30]. The oligonucleotide primers (13 to 15 mers) were synthesized in a Milligen oligonucleotide synthesizer, desalted in a PD-10 Sephadex G-25 M column (Pharmacia) and the first three fractions were dried and then dissolved in 100  $\mu$ l. A 1:100 dilution of the primer was usually used with 0.5  $\mu$ g RNA at a ratio of 1:1 but this ratio and dilution was determined empirically for each primer. The viral RNA and oligonucleotide primer were heated in a 90 °C water bath for 1 min, quickly chilled in an ice/water bath and sequenced with 1.0 unit of AMV reverse transcriptase (Life Sciences), as described [2].

#### *Data analysis*

The neighbour-joining method [29] and various programmes in the SEQ and STATS libraries of the VAX 11-750 of the Research School of Biological Sciences, Australian National University, Canberra were used to analyse the nucleotide and encoded amino acid sequences.

### **Results**

#### *Nucleotide sequences of the NS 1 gene of dengue-2 viruses from Thailand and Sri Lanka*

Figure 1 shows the nucleotide sequence of the NS 1 gene of the dengue-2 16681 virus isolate, as well as the nucleotides that differ from this sequence for four Thai (D, F, G, H) and three Sri Lankan (a, b, c) virus isolates. This figure shows that the nucleotide differences are scattered throughout the entire length of the gene, and there are several positions where the 16681 strain differs from all the others. The numbers of nucleotide differences between each of the viruses shown in Fig. 1 and the published dengue-2 viruses JAM [9], S 1 [12] and NGC [16] are summarized in Table 2. These results indicate that the S 1 virus [12], which is a live attenuated vaccine strain from the Puerto Rican isolate PR 159, is less closely related to the other dengue-2 sequences showing 99–119/1056 nucleotide differences with an average of 106.3/1056 nucleotide differences or 9.8% sequence difference, compared with the range from 27–94/1056 nucleotide differences or 2.6%–8.9% (average 6.1%) difference among the other dengue-2 virus isolates. There are three pairs of viruses which show very few nucleotide

---

**Fig. 1.** Nucleotide sequences of the NS 1 genes of eight dengue-2 viruses, 16681, D, F, G, H, a, b, and c. For complete virus designations see Table 1. Only the changes from the 16681 sequence have been indicated for the other viruses

**Table 2.** Nucleotide and predicted amino acid differences among twelve dengue-2 virus isolates in the gene coding for the non-structural NS1 protein

	Virus isolates <sup>a</sup>										
	16681	D	F	G	H	a	b	c	S1	JAM	NGC
16681	–	64	47	56	67	71	59	61	99	63	34
D	10	–	68	85	27	85	68	71	106	30	53
F	14	21	–	48	74	83	74	82	108	73	59
G	15	22	15	–	89	94	85	93	119	92	72
H	10	4	20	21	–	81	72	77	105	34	54
a	14	14	21	24	14	–	60	73	106	78	57
b	16	14	23	26	14	7	–	49	104	66	30
c	15	11	22	25	11	12	12	–	103	73	48
S1 <sup>b</sup>	16	14	23	24	14	17	18	16	–	98	84
JAM <sup>c</sup>	7	3	18	19	3	11	11	8	11	–	48
NGC <sup>d</sup>	7	5	18	19	5	11	11	10	11	2	–

Numbers above the diagonal represent nucleotide differences out of 1056 nucleotides among the dengue-2 viruses, while those below the diagonal represent amino acid differences out of 352 amino acids

<sup>a</sup> The full names of the viruses which are represented by codes are listed in Table 1

<sup>b</sup> From Hahn et al. [12]

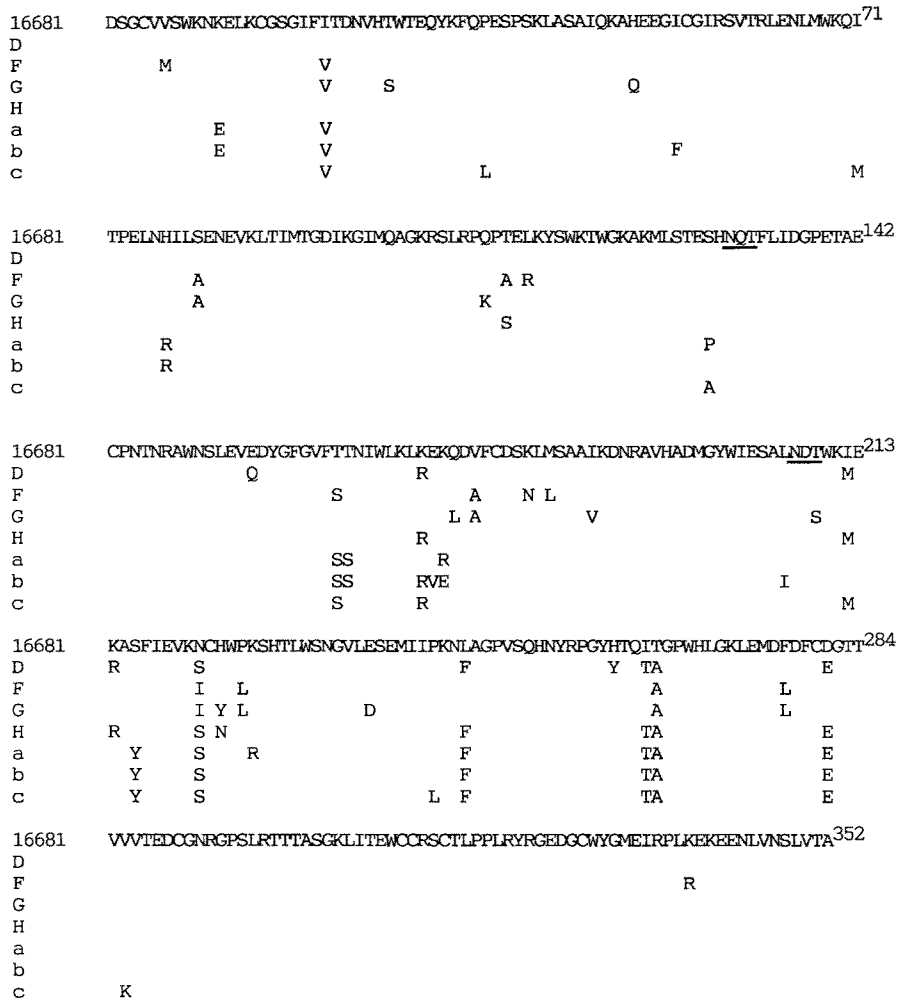
<sup>c</sup> From Deubel et al. [9]

<sup>d</sup> From Irie et al. [16]

differences; (i) D and H, which are both isolates from the 1980 Bangkok epidemic, show 27/1056 differences or 97.4% sequence similarity, (ii) D and JAM which come from geographically separate regions have 30/1056 differences or a 97.2% similarity, and (iii) b and NGC, which represent different geographic regions as well as being isolated 24 years apart are also 97.2% similar.

#### *Amino acid differences in the NS1 protein of the dengue-2 serotype*

The encoded amino acid sequence of the NS1 protein from the 16681 virus isolate is shown in Fig. 2 with only the differences for the other isolates (D, F, G, H, a, b, c). One of the two potential glycosylation sites (NQT at positions 129–131) is found in all of the dengue-2 viruses sequenced so far, whereas the sequence of the second site (NDT at positions 207–209) is in all but one virus isolate (G). This virus isolate has the sequence NDS instead of NDT, thereby still retaining the potential for glycosylation. Various amino acid differences are scattered throughout the protein, especially in region between amino acids 200 to 280. The C-terminal 65 amino acids are closely similar and have only one amino acid difference (R instead of K). Table 2 summarizes the amino acid differences between the Thai and Sri Lankan isolates as well as the published JAM [9], S1 [12], and NGC [16] sequences. These results reveal that the

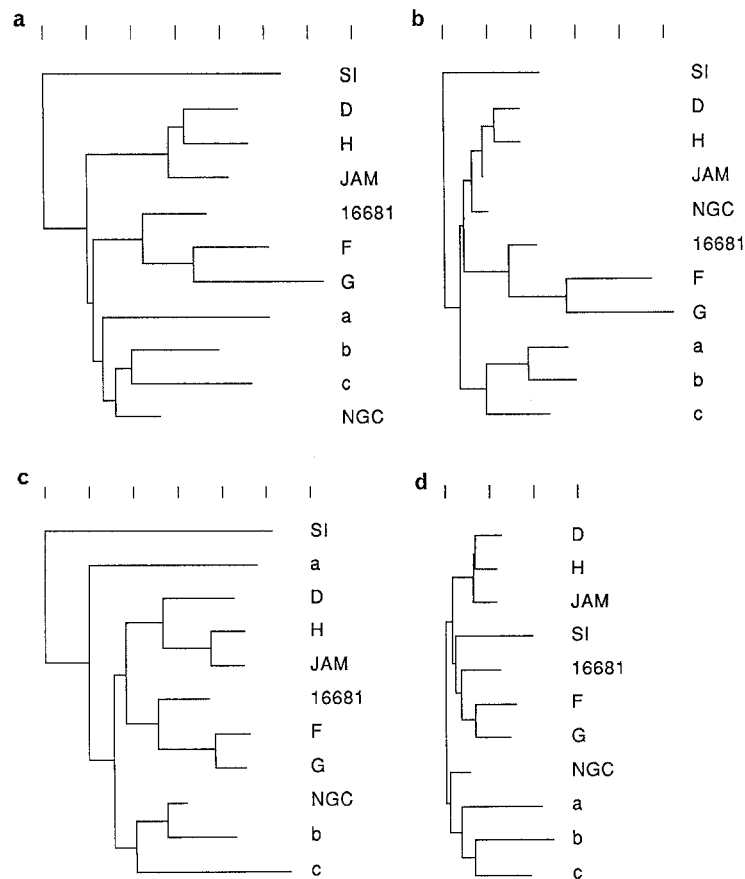


**Fig. 2.** Non-structural NS1 protein sequences derived from translating the eight dengue virus type 2 nucleotide sequences (Fig. 1) are shown in the one letter code. The changes from the 16681 sequence are indicated for the other viruses whose full names are described in Table 1. The two potential glycosylation sites (NXT/S) at positions 129 and 207 are underlined

variability of the amino acids (0.6% to 7.4% difference) is not much less than the nucleotide divergence which ranges from 2.6% to 11.3% difference. The b/NGC pair which has a 97.2% similarity at the nucleotide level is not as closely related at the amino acid level (96.9% similarity) indicating that several of the changes result in amino acid substitutions; whereas the NGC and JAM sequences are more alike at the protein level (2/352 amino acid differences or 99.4% similarity) than at the gene sequence level (48/1056 nucleotide differences or 95.5% similarity).

*Relationships*

The data in Table 2 were converted to percent differences and were used to calculate a similarity network by the neighbour joining (NJ) method [29]. This



**Fig. 3.** Dendrograms illustrating the classifications by the neighbour joining method of **a** the NS 1 nucleotide sequences and **b** the encoded amino acid sequences of the NS 1 glycoproteins from 11 dengue type 2 viruses; and **c** the E nucleotide sequences [5] and **d** the encoded amino acid sequences of the E glycoproteins [5]. The full names of the viruses are shown in Table 1, and scale unit = 1%

is a quick and effective method for calculating phylogenetic trees from distance data, and performs as well as, or better than parsimony methods [17]; it provides more information than a dendrogram calculated by the unweighted pair-group method using arithmetic means (UPGMA) [28] because different rates of evolutionary change are assessed. The resulting dendrograms for the NS 1 genes and proteins (Fig. 3 a and b) are closely similar in structure, even though the ranges of differences are not the same. Firstly, there are three consistent clusters: one contains two Thai isolates (D, H) and the Jamaican isolate (JAM), another contains three Thai isolates (16681 isolated in 1964, and F, G isolated in 1980), and the third cluster contains three Sri Lankan isolates (a, b, c). Secondly, in both classifications, the vaccine strain (S 1) derived from the Puerto Rican isolate PR 159 is distinct, and the position of the NGC isolate is anomalous, grouping with b and c in the nucleotide classification and with the D, H, and JAM cluster in the amino acid classification.



The envelope (E) gene and protein sequences for the eleven dengue-2 viruses [5, 9, 12, 16] were also analysed using the NJ method so that their classifications (Fig. 3 c and d) could be directly compared with those produced for the NS 1 region (Fig. 3 a and b). The classifications for the E and NS 1 genes and encoded proteins are closely similar, although the positions of the a and NGC isolates are different and the inter-cluster relationships vary. The amino acid classifications of the E and NS 1 proteins (Fig. 3 b and d) also show clearly that the NS 1 protein is diverging faster than the E protein. Comparison of the observed nucleotide and amino acid differences between sequences shows that, on average, 19.7% of the nucleotide changes in the NS 1 genes result in amino acid changes, whereas only 15.3% of those in the E genes do. If mutations of the 61 codons specifying amino acids arise at random, about 24% would cause amino acid changes. Thus, it seems that there is selection against amino acid changes in both proteins, especially in the E glycoprotein.

A computer programme, SEQCORR, was used to seek correlations between the nucleotides and amino acids occurring at each position in the aligned sequences and particular characteristics of each of the dengue virus isolates, such as their pathogenicity. Firstly, correlations were sought with the presence of a primary or secondary seroresponse in the patient (Table 1). Secondly, correlations were sought with severity of symptoms, where the DHF and DSS isolates were pooled and compared with the DF isolates (Table 1). These analyses were done by comparing the particular nucleotide or amino acid at each positions and also using standard groupings of amino acids (e.g., hydrophobic, polar, aromatic, etc.). No significant correlations were found. In a further test, the amino acid compositions of the proteins of DHF/DSS isolates and the DF isolates were classified using the Canberra metric and UPGMA clustering [27], and again no correlations were found. A graph plotting the nucleotide and amino acid differences for each pair of sequences showed that comparisons of the DHF/DSS isolates were not unusual and could not be distinguished from the DF isolates. This is unlike that found for influenza virus haemagglutinins, where the source of the virus (human or avian) can be distinguished by the ratio of nucleotide and amino acid differences in the haemagglutinin sequences as this reflects selection for amino acid change in human epidemic influenza [1, 11].

### Discussion

The results presented in this paper show that the NS 1 gene of the eleven dengue-2 virus isolates from patients with various grades of illness, from several countries and different years of isolation varies by 2.6–11.3% at the nucleotide level and 0.6–7.4% at the amino acid level. No correlation was found between year of isolation, serology (1° or 2°) or disease severity and sequence variation. A recent report of the NS 1 sequences from three Malaysian dengue-2 viruses isolated from a dengue fever, DHF and DSS case respectively [10], revealed a 2.3–3.9% nucleotide difference among these three isolates, but only a 0.3–1.4%

amino acid difference. Thus all current data suggest that dengue disease severity is not caused by individual or contiguous features of the NS1 or E protein [5] sequences. Other features of these proteins, such as their nucleic acid interactions or their secondary or tertiary structure, or the activity of other non-structural proteins (e.g., viral polymerase), may define the virulence of particular dengue viruses. However, it should be emphasized that viruses producing disease in humans generally demonstrate the iceberg phenomenon (more sub-clinical and



**Fig. 4.** Dendrogram illustrating the amino acid classification of 40 dengue-2 isolates reported by Rico-Hesse [26], the dengue-2 JAM [9], Thai and Sri Lankan viruses (see Table 1 for details) and the dengue-1 [21] and dengue-4 [20, 39] isolates using the neighbour joining method. The dengue-1 and dengue-4 viruses are the “outgroups” and the dengue-2 viruses divide into two major sets, the upper Group A, containing only West African isolates, and the lower, Groups B, C, and D. Scale units = 5%

mild infections than severe ones), and that there is an inherent heterogeneity of RNA viruses. A particular genetic strain of dengue virus may therefore be isolated from patients showing all grades of disease and the few isolates studied here need not reflect the full diversity of genetic strains prevalent in any country nor can we be certain that they represent the virulent type. The results as such should not be taken to exclude the possibility of significant genetic diversity associated with virulence.

Rico-Hesse [26] sequenced a region of 240 nucleotides spanning the junction of the E and NS 1 genes of 40 dengue-2 isolates, and used the encoded amino acid sequences to classify the isolates using a UPGMA-style of algorithm. To make these comparisons more widely useful, we classified the same data, together with the sequences of the E/NS 1 regions from the JAM [9] isolate and the Thai and Sri Lankan isolates we report here and previously [5]. We classified the resulting 49 viruses by the NJ method to test whether it would provide an estimate of the timing of events in the evolution of dengue-2 viruses; and the homologous sequences for dengue-1 [21] and dengue-4 [20, 39] were used as "outgroups" in the dendrogram.

In the dendrogram (Fig. 4) obtained by the NJ method, the eleven isolates, whose E and NS 1 genes have been fully sequenced, show relationships that closely resemble those calculated from the full NS 1 and E gene sequences (Fig. 3 b and d); an indication of the validity of a classification based on the smaller sample of the genome. This figure also distinguishes between the two 16681 sequences. They differ in the NS 1 portion of the sequence by two nucleotides (UUC to UUU, CCC to CUC, 16681 to Thailand 64 [26], respectively) which result in one amino acid difference (P to L, 16681 to Thailand 64, respectively). This region was also sequenced from cloned cDNA derived 16681 viral RNA (B. L. Herring and J. Blok, unpubl. data) and the results were identical to the 16681 sequence we report here.

The dendrogram places the isolates in the same major groupings as those found by Rico-Hesse [26]. However, there are differences in their horizontal positions in the dendrogram, and this may indicate that they have undergone different amounts of evolutionary change, but there is no obvious correlation between their horizontal positions and the year in which they were isolated. When the four major clusters of isolates in the dendrogram (Fig. 4, A–D) were analysed separately, positive correlations between year of isolation and horizontal position were obtained with all four groups. Data for Senegal 70 (HD 10674), Puerto Rico 69 (PR 159 S 1) and Trinidad 54 (TR 1751) were omitted as they had been experimentally passaged 26, 23, and 57 times respectively [26], whereas all other isolates had been passaged 9 times or less, and "founder effects" during passage are probably responsible for their anomalous positions in the dendrogram. The correlations were statistically significant ( $p < 0.05$ – $0.02$ ) only for Groups B and C. The linear regression expressing the time/position relationships in the Groups gave an estimate of the time of the node joining Groups C and D as 1957.7 (estimated from Group C data), that joining Group

B and Groups C/D as 1953.9 (estimated from Group B data) or 1953.0 (estimated from Group C data), and that joining Group A and Groups B/C/D as 1939.4 (estimated from Group B and C data). The estimates are extrapolations and hence their accuracy diminishes as they increase in length, nonetheless the estimated dates seem sensible since all isolates, except NGC 44, were obtained after the estimated dates of the nodes that connect them.

Group C contains viruses from the Philippines and Taiwan; and Rico-Hesse [26] suggested that the direction of spread of dengue was probably from the Philippines to Taiwan. This is not, however, supported by these data since the two viruses from the Philippines were from 1983 and 1988, and no virus prior to 1981 was sequenced whereas the three Taiwan isolates present in the same cluster were from 1981 (two isolates) and 1987. Although significant correlations with time were obtained for clusters B and C, one must be cautious in using these limited data to predict the origins of dengue outbreaks.

The lack of significant correlation between the time of isolation of Group A viruses and their position in the dendrogram is not unexpected. There are only seven isolates in Group A, one had been repeatedly passages and all, except one, of the remainder were isolated in the years 1980 or 1981. However, detailed comparisons of the sequences also suggest that Group A isolates vary from one another in ways that differ from those found with the other isolates. Group A isolates differ from one another by approximately equal numbers of transitions (i.e., changes from purine to purine, G to A, or pyrimidine to pyrimidine, C to U) and transversions (i.e., changes between purine and pyrimidine) with an average of  $2.2 \pm 0.8$  and  $1.3 \pm 0.9$ , respectively; whereas isolates of the other three groups have more transition than transversion differences (an average of  $6.5 \pm 2.4$  and  $1.2 \pm 0.7$ , respectively). These differences suggest that the evolutionary processes operating in the two virus populations are not the same. There is evidence from many studies [19] that transitions occur more frequently, especially in third silent codon positions, during the early evolutionary divergence of nucleotide sequence. Transversions, on the other hand, occur much less frequently but when they occur, especially in second codon positions, they usually cause amino acid changes in the encoded proteins.

The Group A viruses were all from West African countries and six out of the seven were isolated from mosquitoes, whereas the other (Senegal 70) was from human serum, as were the other 42 viruses used in this classification. Rico-Hesse [26] suggested that the Group A isolates represent a forest cycle of dengue in West Africa and that "epidemic" dengue viruses from the other areas have evolved independently. In her study, nine viruses were isolated from West African countries (three from human sera and six from mosquitoes) of which seven were in Group A while the others were in Group D. Too few dengue viruses isolated from human sera in West African countries have been sequenced and more isolates must be examined to allow a significant assessment to be made of the importance of transition and transversion rates among these viruses. Nonetheless the observed differences indicate that it is not safe to

extrapolate from the correlations observed with Groups B, C, and D to the timing of the gap between the West African Group A isolates and the epidemic dengue-2 isolates.

These results provide a database for dengue virus epidemiology and systematics, and may contribute significantly to the understanding and control of dengue.

**Note added in proof.** The nucleotide sequences reported in this paper have been submitted to GenBank and are assigned accession numbers M58486–M58493 for 16681, D, F, G, H, a, b, c, respectively.

### Acknowledgements

We would like to thank Drs A. Nisalak and E. Henchal for recovering some of the Sri Lankan isolates during transport and Dr. B. Gorman for providing the dengue virus stocks for some of the isolates. This research was supported by the National Health and Medical Research Council of Australia and the World Health Organization as part of its Programme of Vaccine Development.

### References

1. Air GM, Gibbs AJ, Laver WG, Webster RG (1990) Evolutionary changes in influenza B are not primarily governed by antibody selection. *Proc Natl Acad Sci USA* 87: 3884–3888
2. Blok J, Air GM (1980) Comparative nucleotide sequences at the 3' end of the neuraminidase gene from eleven influenza type A viruses. *Virology* 107: 50–60
3. Blok J, Henchal EA, Gorman BM (1984) Comparison of dengue viruses and some other flaviviruses by cDNA-RNA hybridization analysis and detection of a close relationship between dengue virus serotype 2 and Edge Hill virus. *J Gen Virol* 65: 2173–2181
4. Blok J, Kay BH, Hall RA, Gorman BM (1988) Isolation and characterization of dengue viruses serotype 1 from an epidemic in northern Queensland, Australia. *Arch Virol* 100: 213–220
5. Blok J, Samuel S, Gibbs AJ, Vitarana UT (1989) Variation of the nucleotide and encoded amino acid sequences of the envelope gene from eight dengue-2 viruses. *Arch Virol* 105: 39–53
6. Brandt WE, Cardiff RD, Russell PK (1970) Dengue virions and antigens in brain and serum of infected mice. *J Virol* 6: 500–506
7. Brandt WE, Chiewslip D, Harris DL, Russell PK (1970) Partial purification and characterization of a dengue virus soluble complement fixing antigen. *J Immunol* 105: 1565–1568
8. Cardiff RD, Lund JL (1976) Distribution of dengue-2 antigens by electron immunocytochemistry. *Infect Immun* 13: 1699–1709
9. Deubel V, Kinney RM, Trent DW (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
10. Fong MY, Koh CL, Samuel S, Pang T, Lam SK (1990) Nucleotide sequences of the nonstructural protein NS1 gene of three dengue-2 viruses, M1, M2, and M3, isolated in Malaysia from patients with dengue haemorrhagic fever, dengue shock syndrome and dengue fever, respectively. *Nucleic Acids Res* 18: 1642

11. Gibbs AJ, Air GM, Laver WG (1982) Analysis of variation among haemagglutinin genes of influenza A viruses. In: Mackenzie JS (ed) *Viral diseases in South-East Asia and the Western Pacific*. Academic Press, New York, pp 546–549
12. Hahn YS, Galler R, Hunkapiller T, Dalrymple JM, Strauss JH, Strauss EG (1988) Nucleotide sequence of dengue 2 RNA and comparison of the encoded proteins with those of other flaviviruses. *Virology* 162: 167–180
13. Halstead SB, Simasthien P (1970) Observations related to the pathogenesis of dengue hemorrhagic fever II. Antigenic and biologic properties of dengue viruses and their association with disease response in the host. *Yale J Biol Med* 42: 276–292
14. Henchal EA, Henchal LS, Schlesinger JJ (1988) Synergistic interactions of anti-NS 1 monoclonal antibodies protect passively immunized mice from lethal challenge with dengue 2 virus. *J Gen Virol* 69: 2101–2107
15. Igarashi A (1978) Isolation of a Singh's *Aedes albopictus* cell clone sensitive to dengue and Chikungunya viruses. *J Gen Virol* 40: 531–544
16. Irie K, Mohan PM, Sasaguri Y, Putnak R, Padmanabhan R (1989) Sequence analysis of cloned dengue virus type 2 genome (New Guinea-C strain). *Gene* 75: 197–211
17. Jin L, Nei M (1990) Limitations of the evolutionary parsimony method of phylogenetic analysis. *Mol Biol Evol* 7: 82–102
18. Kerschner JAH, Vorndam AV, Monath TP, Trent DW (1986) Genetic and epidemiological studies of dengue type 2 viruses by hybridization using synthetic deoxyoligonucleotides as probes. *J Gen Virol* 67: 2645–2661
19. Li W-H, Wu C-I, Luo C-C (1984) Non randomness of point mutations is reflected by nucleotide substitutions in pseudogenes and its evolutionary implications. *J Mol Evol* 21: 58–71
20. Mackow E, Makino Y, Zhao B, Zhang Y-M, Markoff L, Buckler-White A, Guiler M, Chanock R, Lai C-J (1987) The nucleotide sequence of dengue type 4 virus: analysis of genes coding for nonstructural proteins. *Virology* 159: 217–228
21. Mason PW, McAda PC, Mason TL, Fournier MJ (1987) Sequence of the dengue-1 genome in the region encoding the three structural proteins and the major nonstructural protein NS 1. *Virology* 161: 262–267
22. Monath TP, Wands JR, Hill LJ, Brown NV, Marciniak RA, Wong MA, Gentry MK, Burke DS, Grant JA, Trent DW (1986) Geographic classification of dengue-2 virus strains by antigen signature analysis. *Virology* 154: 313–324
23. Putnak JR, Charles PC, Padmanabhan R, Irie K, Hoke CH, Burke DS (1988) Functional and antigenic domains of the dengue-2 virus nonstructural glycoprotein NS 1. *Virology* 163: 93–103
24. Repik PM, Dalrymple JM, Brandt WE, McCown JM, Russell PK (1983) RNA fingerprinting as a method for distinguishing dengue 1 virus strains. *Am J Trop Med Hyg* 32: 577–589
25. Rice CM, Strauss EG, Strauss JH (1986) Structure of the flavivirus genome. In: Schlesinger S, Schlesinger M (eds) *The Togaviridae and Flaviviridae*. Plenum, New York, pp 279–327
26. Rico-Hesse R (1990) Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. *Virology* 174: 479–493
27. Rohlf FJ (1989) *NTSYS-pc: numerical taxonomy and multivariate analysis system*. Exeter, New York
28. Rohlf FJ, Wooten MC (1988) Evaluation of the restricted maximum-likelihood method for estimating phylogenetic trees using simulated allele-frequency data. *Evolution* 42: 581–595
29. Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406–425

30. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463–5467
31. Sangkhawibha N (1982) Viral diseases in Thailand – a national overview. In: Mackenzie JS (ed) *Viral diseases in South-East Asia and the Western Pacific*. Academic Press, New York, pp 217–223
32. Schlesinger JJ, Brandiss MW, Walsh EE (1987) Protection of mice against dengue-2 virus encephalitis by immunization with dengue-2 non-structural glycoprotein NS 1. *J Gen Virol* 68: 853–857
33. Schlesinger RW (1977) *Dengue viruses*. Springer, Wien New York [Gard S, Hallauer C (eds) *Virology monographs*, vol 16]
34. Trent DW, Grant JA, Rosen L, Monath TP (1983) Genetic variation among dengue 2 viruses of different geographic origin. *Virology* 128: 271–284
35. Trent DW, Grant JA, Monath TP, Manske CL, Corina M, Fox GE (1989) Genetic variation and microevolution of dengue 2 virus in Southeast Asia. *Virology* 172: 523–535
36. Walker PJ, Henchal EA, Blok J, Repik PM, Henchal LS, Burke DS, Robbins SJ, Gorman BM (1988) Variation in dengue type 2 viruses isolated in Bangkok during 1980. *J Gen Virol* 69: 591–602
37. Westaway EG, Brinton MA, Gaidamovich SYa, Horzinek MC, Igarashi A, Kääriäinen L, Lvov DK, Porterfield JS, Russell PK, Trent DW (1985) *Flaviviridae*. *Intervirology* 24: 183–192
38. World Health Organization (1975) *Technical guides for diagnosis, treatment, surveillance, prevention and control of dengue hemorrhagic fever*. WHO, Geneva
39. Zhao B, Mackow E, Buckler-White A, Markoff L, Chanock RM, Lai C-J, Makino Y (1986) Cloning full-length dengue 4 viral DNA sequences: analysis of genes coding for structural proteins. *Virology* 155: 77–88

Authors' address: Dr. J. Blok, Sir Albert Sakzewski Virus Research Laboratory, Royal Children's Hospital, Herston Road, Brisbane, Qld. 4029, Australia.

Received September 7, 1990