



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

## Comparison of a Dengue-2 Virus and Its Candidate Vaccine Derivative: Sequence Relationships with the Flaviviruses and Other Viruses<sup>1</sup>

J. BLOK,<sup>2</sup> S. M. McWILLIAM,<sup>3</sup> H. C. BUTLER, A. J. GIBBS,\* G. WEILLER,\* B. L. HERRING, A. C. HEMSLEY, J. G. AASKOV,† S. YOKSAN,‡ AND N. BHAMARAPRAVATI#

*Sir Albert Sakzewski Virus Research Centre, Royal Children's Hospital, Herston Rd, Brisbane, Queensland 4029, Australia;*

*\*Molecular Evolution and Systematics Group, Research School of Biological Sciences, Australian National University, Canberra City, ACT 2601, Australia; †Queensland University of Technology, George St., Brisbane, Queensland 4000, Australia; and ‡Center for Vaccine Development, Mahidol University, Nakhon Pathom, Thailand*

Received September 10, 1991; accepted December 2, 1991

A comparison of the sequence of the dengue-2 16681 virus with that of the candidate vaccine strain (16681-PDK53) derived from it identified 53 of the 10,723 nucleotides which differed between the strains. Nucleotide changes occurred in genes coding for all virion and nonvirion proteins, and in the 5' and 3' untranslated regions. Twenty-seven of the nucleotide changes resulted in amino acid alterations. The greatest amino acid sequence differences in the virion proteins occurred in prM (2.20%; 2/91 amino acids) followed by the M protein (1.33%; 1/75 amino acids), the C protein (0.88%; 1/114 amino acid), and the E protein (0.61%; 3/495 amino acids). Differences in the amino acid sequence of nonvirion proteins ranged from 1.51% (6/398 amino acids) in NS4 to 0.33% (3/900 amino acids) in NS5. The encoded protein sequences of 16681-PDK53 were also compared with the published sequences of other flaviviruses to obtain a detailed classification of 17 flaviviruses using the neighbor-joining tree method. The analyses of the sequence data produced dendrograms which supported the traditional groupings based on serological evidence, and they suggested that the flaviviruses have evolved by divergent mutational change and there was no evidence of genetic recombination between members of the group. Comparisons of the sequences of the flavivirus polymerase and helicase-like proteins (NS5 and NS3, respectively) with those from other viruses yielded a classification of the flaviviruses indicating that the primary division of the flaviviruses was between those transmitted by mosquitoes and those transmitted by ticks. © 1992 Academic Press, Inc.

### INTRODUCTION

Dengue viruses, of which there are four serotypes, belong to the *Flaviviridae*, a family of closely related viruses transmitted by mosquitoes and ticks (Westaway *et al.*, 1985). These viruses produce enveloped particles containing a single positive stranded RNA genome of ca. 11 kb, which codes for the three virion proteins [Capsid (C), prM/Matrix (prM/M), and Envelope (E)] and seven nonvirion proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5); see the review in Rice *et al.* (1986).

All four serotypes of the dengue virus have been isolated from patients with classical dengue fever as well as from patients with the more severe form of the disease, dengue hemorrhagic fever/dengue shock syndrome (Schlesinger, 1977). Of the millions of cases of

dengue each year, most are uncomplicated dengue fever which is a mild febrile illness characterized by headache, joint pain, fever, and rash. However, some patients develop a thrombocytopenia, increased vascular permeability and, in severe cases, hypovolemic shock which may lead to death if untreated. The risk of developing severe dengue infections (DHF/DSS) is greater in patients with anamnestic infections, particularly where the serotype causing the anamnestic infection differs from that which caused the earlier infection (Halstead, 1981). It has been reported that some cases of DHF/DSS could be due to virulent strains of dengue viruses (Rosen, 1977).

In endemic areas, immunity against the four dengue serotypes can be demonstrated in the majority of the adult population, and therefore protective immunity may be gained by immunization with appropriate dengue vaccines. An ideal vaccine would (i) be effective against all four dengue virus serotypes, (ii) induce >95% seroconversion, (iii) produce lifelong immunity, and (iv) be administered to children aged 6 months to 1 year. The dengue vaccine development program at the Mahidol University in Bangkok has attempted to attenuate all four dengue serotypes by passage *in vitro* in cell

<sup>1</sup> Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. M85258 and M85259

<sup>2</sup> To whom correspondence and reprint requests should be addressed.

<sup>3</sup> Present address: CSIRO Division of Tropical Animal Production, Long Pocket Laboratories, Brisbane, Queensland, 4068, Australia.

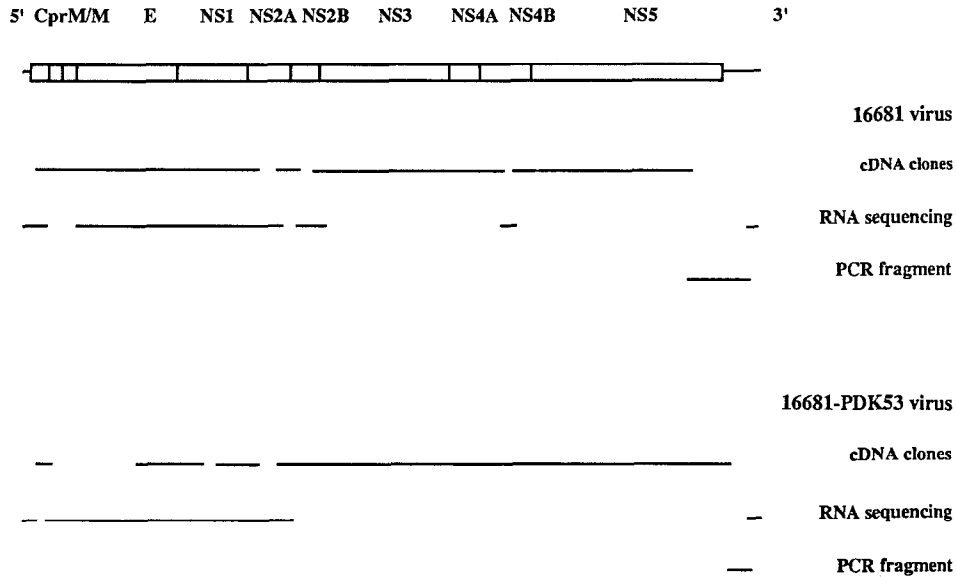


FIG. 1. Schematic diagram of the dengue-2 16681 and 16681-PDK53 RNA genomes showing the regions which were sequenced from (i) cDNA clones (cloned fragments were represented by more than one clone in most instances and were sequenced from both strands), (ii) viral RNA genome using specific oligonucleotide primers, and (iii) fragments produced by the polymerase chain reaction (344 bp for both 16681 and 16681-PDK53, and another 550 bp fragment for 16681). The dideoxy sequencing technique (Sanger *et al.*, 1977) was used in all cases; see text for details.

culture. A candidate live dengue-2 virus vaccine has been produced from the dengue-2 16681 virus isolated from a fatal case in Thailand (Halstead and Simasthien, 1970) by 53 serial passages in primary dog kidney (PDK) cells. This dengue-2 16681-PDK53 candidate vaccine virus appears to be biologically attenuated in several characters (Yoksan *et al.*, 1986), and Phase I trials with human volunteers have been completed (Bhamarapravati *et al.*, 1987). These trials indicated that the vaccine produced a good immune response and no recipient showed any clinical symptoms or severe side effects. It was therefore decided to sequence the genome of these dengue-2 viruses (16681 and 16681-PDK53) in order to define the molecular differences between them. This is a necessary step toward identifying the molecular basis of virulence.

We report the comparison of the nucleotide sequences of the dengue-2 16681 and 16681-PDK53 viruses with the sequences of other flaviviruses: dengue-1, -2, -3, and -4, and yellow fever, West Nile, Kunjin, Japanese encephalitis, Murray Valley encephalitis, St. Louis encephalitis, and tick-borne encephalitis.

## MATERIALS AND METHODS

### Virus isolates

The dengue-2 16681 virus strain was isolated from a fatal case of DSS in Thailand in 1964 (Halstead and Simasthien, 1970) and passaged twice in *Toxorhynchites ambionensis* mosquitoes and once in primary green monkey kidney cells to produce the "parent virus" for attenuation studies, Yoksan *et al.* (1986). The dengue-2 16681-PDK53 virus was obtained by serial passage of the "parent virus" in PDK cells. The 16681 virus was passaged weekly in PDK cells for 53 passages and at this point it appeared to be attenuated as judged by biological markers such as plaque size, temperature sensitivity, and mouse and monkey neurovirulence (Yoksan *et al.*, 1986). The seed stocks for both 16681 and 16681-PDK53 viruses were grown for one passage in *Aedes albopictus* cells (clone C6/36) (Igashii, 1978), before large scale propagation in C6/36 cells for cloning and sequencing experiments.

### Growth of viruses and purification

C6/36 *Aedes albopictus* cells were grown at 32° in RPMI 1640 medium containing 10% fetal calf serum,

FIG. 2. Nucleotide and encoded amino acid sequence of dengue-2 16681-PDK53 virus. This is a consensus sequence based on the sequencing strategies shown in Fig. 1. The location of the N- and C-termini of the proteins is based on Kunjin virus (Coia *et al.*, 1988; Speight *et al.*, 1988), and the potential glycosylation sites of the virion and NS1 proteins are circled. The 3' terminal 14 nucleotides have not been deduced but are assumed to be the same as those found in West Nile virus (Wengler and Wengler, 1981) since the 14-mer complementary to the West Nile sequence hybridized well to the dengue-2 RNA during cDNA synthesis and sequencing experiments.

a

AGUUGUUAUGUCUACGUGGACCCACAAAGACAGAUUCUUGAGGGAGCUGAGCUCAAUGUAGUUCUAAACUGUUUUUUGAUUAGAGACGAGAUUCUGAUG 99  
M 1

→C  
AAUGACCAACCGGAAAGGGCAAAAACACGCCAUUCAUUGCUGAAACCGGAGAGAAACCGGUGUCGACUGGCAACAGCUGACAAAGAGAUUCUCA 198  
N D Q R K E A K N T P F N M L K R E R N R V S T V Q Q L T K R F S 34

CUUGGAAUGUCUCAGGGACGAGGACCAUUAACUUAACUUGGCCUGGGUGGUUCCUUCGUUUCCUAAUCCACCAACAGCAGGGAUUAUGAAG 297  
L G M L Q G R G P L K L Y M A L V A F L R F L T I P P T A G I L K 67

AGAUGGGGAAAUUUAAAAUCCAAAGCUAUUAUUGUUGAGAGGGUUCAGGAAAGAGAUGGAAAGGAGUCGUAACAUUUGAAUAGGAGACGAGA 396  
R W G T I K K S K A I N V L R G F R K E I G R M L N I L N R R R R 100

UCUGCAGGCAUGAUUAUGCUGAUUCCAAACAGUAGGGCGUCCAUUAACCAACGUAACGGAGAACCACACAUAGUCGACAGACAGAAAGAGAAA 495  
S A G M I I M L I P T V M A F H L T T R N G E P H M I V S R Q E K 133

GGGAAAGUCUUCGUUUUAAACAGAGGUGGGGUAACUAGUGUACCCUACUGGCCAUGGACCUUGGUAUUGUGUGAAGACACAAUACGUAACAAG 594  
G K S L L F L K T E V G V N M C T L M A M D L G E L C E D T I T Y K 166

UGUCCCUUCUCAGGCGAAUAGGCCAGAAAGCAUAGACUGUUGGCAACUCUACUGCCAGUGGUAACUUAUGGACGUGUACCCACCAUGGGAGAA 693  
C P L L R Q N E P E D I D C W C N S T S T W V T Y G T C T T M G E 199

→M  
CAUAGAAGAGAAAAAGUACAGUGGCCACUCGUUCCACAUUGGGUAGGGGACGAGACACGAAACUAGGUAUCAGAAAGGGCCUGGAAA 792  
H R R E K R S V A L V P H M G L E T R T E T W M S S E G A W K 232

CAUGUCCAGAGAAUUGAAACUUGGACUUGAGACAUCCAGGCUUACCAUGAGGCAUCCUGGCAUACCAUAGGAAACGACACAUUCCAAAGA 891  
H V Q R I E T W I L R H P G F T M M A A I L A Y T I G T T H F Q R 265

GCUCGAAUUUAUUCUGACAGCUGUCACUCCUCAAUGACAUUGCCUUGCAUAGGAAUUGUCAAUAGAGACUUUGGAAAGGGGUUACAGGAGGA 990  
A L I L L L T A L V T M R C I G M S N R D F V E G V S W K 298

AGCUGGUUGACAUAGUCUUAAGAAACUAGGAGCUGUGACGACGAGUAGGCAAAAACAAACCAUUGGAAUUGAUCUGAUAAGAAACAGAAAGCCAAA 1089  
S W V D I V L E H G S C V T T M A K N K P T L D F E L I K T E A K 331

CAGCCUGCCACCUAAGGAAGUACUGUAUAGGCAAGCUAACCAACACAAACAGAAUUCGUGCCCAACACAAAGGGAAACCCAGCCUAAAUGAA 1188  
Q P A T L R K Y C I E A K L T N T T E S R C P T Q G E P S L N E 364

GAGCAGGACAAAGGUUCGUCGCAACACUCCAUUGGAGACAGAGGAAUGGAAUUGGACUUAUUGGAAAGGGAGGCAUUGGACCUUGUCU 1287  
E Q D K R F V C K H S M V D R G W G N G C G L F G K G G I V T C A 397

AUGUUCAGAUCAAAAGAAACUAGGAAAGGAAUUGUGCAACCAAAAACUUGGAAUACCAUUGGAAUACACCUCACUGGGGAAAGAGCAUUGCA 1386  
M F R C K K N M E G K V V Q P E N L E Y T I V I T P H S G E H A 430

GUCGGAAUAGACACAGAAAAACUAGGCAAGAAUCAAUUAACACACAGAGUUCUACACAGAGCAGAAUUGACAGGUUAUGGCAUUCACAAUG 1485  
V G N D T G K H G K E I K I T P Q S S I T E A E L T G Y G T I T M 463

GAGUCUCUCCAGAAAGCGGCUAGACUCAAUGAGAUUGGUGUGCAGUAGGAAUUAAGCUUGGCUUGGACAGGCAUUGGUUCCUAGACCUG 1584  
E C S P R T G L D F N E I V L L Q M E N K A W L V H R Q W F L D L 496

CCGUUACCAUGGUGCCCGGACCCACACACAGGCUAAUUGGAAUACAGAAAGACAUUGGACUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 1683  
P L P W L P G A D T Q G S N W I Q K E T L V T F K N P H A K K Q D 529

GUUGUUGUUUAGGAAUCCAGAGAGGGCCACACAGCAGCUCUACAGGGCCACAGAAUCCAAUUGUACUACAGGAAACUUCUUCACAGGACAU 1782  
V V V L G S Q E G A M H T A L T C A T E I Q M S S G N L L T C H 562

CUCAAUGGACGAGGAAUAGGAAAGCAGCUCUCAAAGGAAUUGUACUUCUUAUGGCAAGAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 1881  
L K C R L R M D K L Q L K G M S Y S H C T G K F K V V K E I A E T 595

CAACAGGAAACAAUAGUUAUAGAGUCCAAUUAAGAGGGGACGGCUCUCCUAGGAAUCCUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 1980  
Q H G T I V I R V Q Y E G D G S P C K I P F E I M D L E K R H V L 628

GGUCGCGAUUAUAGUACUAAUUGGACAGAAAGAAUAGCCAGUACAAUAGAAAGCAGAAUCCUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 2079  
R L I T V T E K D S P V N I E A E P P P G D S Y I I I G 661

GUAGACCGGAAACUGAAGCUCACUGGUUUAAGAAAGGAAUUCUUAUUGGCAAAUUGUAGACAAUAGAGGGGGGAGAGAAUUGGCAUUA 2178  
V E P G Q L K L N W F K K G S S I G Q M F E T T M R G A K R M A I 694

UUAGGUGACAGCCUGGAAUUGGAAUUGGAAUUGGAAUUGGAAUUGGAAUUGGAAUUGGAAUUGGAAUUGGAAUUGGAAUUGGAAUUGGAAUUG 2277  
L G D T A W D F G S L G G V F T S I G K A L H Q V F G A I Y G A A 727

UUCAGUGGGUUUUAUGGACUUA 2376  
F S G V S W T M K I L I G V I I T W I G M N S R S T S L S V T L V 760

→NS1  
UUGGUGGAAUUGGACUGUUAUUGGAGUUAUUGGAGCAGGCGAUGUUGGUGUGGUGGAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 2475  
L V G I T L Y L G V M V Q A D S G C V V S W K N K E L K C G S G 793

AUUUUAUCACAGACACGUGCACACUAGGACAGAAACAAUACAAUCCAGAAUCCCUUCAAACUAGCUUACGCUUCCAGAAAGCCCAUCAA 2574  
I F I T D N V H T W T E Q Y K F Q P E S P S K L A S A I Q K A H E 826

GAGGACAUUUGGAAUCCGUCAGUAAACAGACUGGAGAAUUGUAGGAAACAAUUAACACAGAAUUGAUAUCACAUUUAUUAUUAUUAUUA 2673  
E D I C G I R S V T R L E N L M W K Q I T P E L N H I L S E N E V 895

AAGUUAACUUAUUAUGACAGGACACAAAGAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 2772  
K L T I M T G D I K G I M Q A G K R S L R P Q P T E L K Y S W K T 892

UGGGCAAGGAAUUGCUCUACAGAGUCUUA 2871  
W G K A K M L S T E S H N Q T F F I D G P E T A E C P N T N R A W 925

AAUUCGUUGAAGUAGAGAUUAGGCUUUGGAGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 2970  
N S L E V E D Y G F G V F T T N I W L K L K E K Q D V F C D S K L 958

AUGCAGCGCCAAUUAAGAAACAGAGCCGUCUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3069  
M S A A I K L D N R A V H A D M G Y W I E S A L N D T W K I E K A S 991

UUCAUUAAGUAAAAACUGCCACUGGCAAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3168  
F I E V K N C H W P K S H T L W S N G V L E S E M I I P K N L A G 1024

CCAGUGUCUUAACCAACUUAUAGACAGGCUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3267  
P V S K H R P P G Y H T Q I T G P W H L G K L E M D F D F C D G 1057

ACAACAGUGUAGUGACUGAGGACUGCGGAAUAGAGGACCCUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3366  
T T V V V T E D C G N R G P S L R T T T A S G K L I T E W C C R S 1090

UGCACAUUACCCCGCUAAGAUACAGAGGUGAGGAGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUG 3465  
C T L P L R Y R G E D G C W Y G M E I R P L K E K E E N L V N S 1123

→NS2A  
UUGGUCACAGCUGGACUAGGACAGGACAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3564  
L V T A G H G Q V D N F S L G V L G M A L F L E E M L R T R V G T 1156

AAACAUGAAUUA 3663  
K H A I L L V A V S F V T L I I G N R S F R D L G R V M V M V G C A 1189

Fig. 2—Continued



C

UUGGCAAGUAUUGCCUAGUCUCUGCGUGACUCAAGUUAUGAUGAGGACUACUAGGCGUCUGUGAGGCUUUAACCUAGUACCGGGCCC	7425
L G Q V M L L V L C V T Q V L M M R T T W A L C E A L T L A T G P	2443
AUCUCCCAUUGCGGAGGAAUCCAGGAGUUUUGGAACACUACCAUUGCGGUGCAAUGGCUAAUUAUAGAGGAGUUAUCUUGGCGGAGCU	7524
I S T L S E G N P G R F W N T T I A V S M A N I F R G S Y L A G A	2476
GGACUUCUUCUUAUUGAAGAACAACAACAACAGAAAGGUAACUGGUAACUAGGAGAGCGCUGGAGAGAAUUGGAAAGCCGAUUGAAC	7623
G L L F S I M K N T T N T R R V T G N I G E T L G E K W K S R L N	2509
GC AUUGGAAAAGUGAAUCCAGAUCAAGAAAAGUGGAAUCCAGGAUGGAAUAGAACCUUAGCAAAGAAGGCAUUAUAAAAGGAGAAACGGAC	7722
A L G K S E F Q I Y K K S G I Q E V D R T L A K E G I K R G E T D	2542
CAUCACGUGUGCGGAGGCGCAGCAAACUGAGAUGGUUCGUGAGAAACAUGGUCACACCAGGAAAGUAGUGGACCCUGGUGUGGCGAGA	7821
T Y G M P S R G S A K L R W F E R N M V T P E G K V V D L G C G R	2575
GGAGGCGUGUCAUUAUUGUGGAGGACUAAAGAAUGUAGAGAAUGCAAAGGCCUAAACAAAAGGAGGACCGAGACGAGAAACCCAUCCCAUGUCA	7920
G G W S Y Y C G G L K N V R E V K G L T K G G P G H E E P I P M S	2608
ACAUAUGGUGGAAUCUAGUGCCUUCUCAAAGUGGAGUAGCGUUUUCUUAUCCGCGGAGAAAGUGACACAUUAUUGUGACAUAGGGGAGUCA	8019
N T Y G M P D V D L G S G T R N I G I E S E I P N L D I I G C R I E S	2641
UCACCAAUCCACAGUGGAGGAGCAGCAACACUCAGAGUCCUUAACUAGUAGAAAUUGUUGAACAAACACUCUAAUUAUUGCAUAAAGGUUCUC	8118
S P N P T V E A G R T L R V L N L V E N W L N N N T Q F C I K V L	2674
AACCCAUUAUUGCCUCAGUCAVAGAAAAAUGGAAGCACUCAAAGGAAUUAUGGAGGAGCCUAGUGAGGAAUCCACUCACGAAACCCACACAU	8217
N H A V S R G S I E K M E A L Q R K Y G G A L V R N P L S R N S R	2707
GAGAUGUACUGGUAUCCAAUGCUCCGGAACAUGUGUCAUCAGUGAACAUUAUUAAGGAUGUAGUACACAGAUUAUUAUAGAGAUACAAGAAA	8316
E M Y W V S N A S G N I V S S V N M I S R M L I N R F T M R Y K K	2740
GCCAUUCAGCGCGGAGUUGACCCGGAAGCGGAACCCGUAACUGGAAUGAUGAGAUACCAAACCUAGAUUAUUAUUGGAAAAGAAUAGAA	8415
A T Y E P D V D L G S G T R N I G I E S E I P N L D I I G C R I E S	2773
AAAAUAAAGCAAGAGCAUGAAACAUAUGGCACUAUGACCAAGACCCCAUACAAAACGUGGCAUACCAUGGUGCUAUGAAACAAAACAGACUGGA	8514
K I K Q E H E T S W H Y D Q D H P Y K T W A Y H G S Y E T K Q T G	2806
UCAGCAUCAUUGGCAACGAGGUGUUCAGGCGUGCACAACCUUGGAGCGUUGUCCCAUGGUGACACAGUUGGCAUAGACAGACAGACUCCA	8613
A S S M V N G V F R L L T K P V D V V P M V T Q M A M T T P P	2839
UUUGGCAACAGCGCGUUUUAAGAGAAAGUGGACAGGAAACCAAGAACCGAAAGAGGACGAAAGAAACUUAUAGAAAUAACAGCAGAGUGGCUU	8712
F G Q Q R V F K E K V D T R T Q E P K E G T K K L M K I T A E W L	2872
UGGAAAGAAUUGGAAAGAAAAGACACCCAGGAUGCTCAGAGAAAGAAUUAACAAGAAAGUGAGAAAGCAUAGCAGCCUUGGGGCCAUUAUCAC	8811
W K E L G K K K T P R M C T R E E F T R K V R S N A A L G A I F T	2905
GAUGAGAAACAUGGAAAGUGGCGACGUGAGGCGUGUAGAAUAGUAGGUUUUGGAGCGUGUAGCAAGGAAAGAAUCCUACUUGAAGGAAAGUGU	8910
D E N K W K S A R E A V E D S R F W E L V D K E R N L H L E G K C	2938
GAACAUGUGUACUACAUAUGGAAAAGAGAGAAGCGUAGGGAAUUCGCGAAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG	9009
E T C V Y N I M G K R E K K L G E F G G K A K G S R A I W Y M W L G	2971
GCACCUUCUUAAGAUUGAAGCCUAGCAUUCUUAUUAUGAAGAUACUGGUCUCCAGAGAGAACUCCUGAGUGGAGUGGAGGAGAGGAGGAGGCGCAC	9108
A R P L E F E A L G F L N E D H W F S R E N S L S G V E G E G L H	3004
AAGCUAGUUAACUUAAGAGAGCGUGAGCAAGAAAGGGAGGCAUUGUCCGAGUACACCCGAGGAGGAAUACAAAGAAUACACACUAGAAGAC	9207
K L G Y I L R D V S K K E G G A M Y A D D T A G W D T R I T L E D	3037
CUAAAAUAGAAAGUUAACAACCAACUAGGAGGAGAAACAAAGAAACUAGCGGACCCAUUUUAACUUAACUUAACUUAACUUAACUUAACUUAAC	9306
L K N E E M V T N H M E G E H K K L A E A I F K L T Y Q N K V V R	3070
GUGCAAAGCAACCAAGGACAGUUAUGGACAUCAUUCGAGAAGAGCAAAAGGAGUUGGAGGAAUUGGAGGAAUUGGAGGAAUUGGAGGAAUUGG	9405
V Q R P T P R G T V H D I I S R R D Q R G S G Q V G T Y G L N T F	3103
ACCAUUAUGGAAAGCCCAUUAUCAGACAGUAGGAGGAGAGGAGUUAUUAAGCAUUCAGCACCUAACAUAUCAGAAAGAAUUGCUGUGCAAAC	9504
T N M E A Q L I R Q H E G E G V F K S I Q H L T I T E E I A V Q N	3136
UGUUAAGCAAGAGGCGCGAAGGUUAUCAAGAAUGGCCAUCAGUGGAGUUAUGGAGGAAUUGGAGGAAUUGGAGGAAUUGGAGGAAUUGGAGGAAUUGG	9603
W L A R V G R E R L S R M A I S G D D C V V K P L D D R L P S A L	3169
ACAGCUCUUAAGACAUUGGAAAGAAUUGGAAAGCAUACAACAAGGAAACCUUACAGAGGAAUUGGAAUUGGAAUUGGAAUUGGAAUUGGAAUUGGAAUUGG	9702
T A L N D M G K I R K D I Q Q W E P S R G W N D W T Q V P F C S H	3202
CAUUUUGCUGCGGAGACCGGCGUUCUAGG	9801
H F H E L I M K D G R V L V V P C R N Q D E L I G R A R I S Q G A	3235
GGUGUGCUUUGCGGAGACCGGCGUUGGAGG	9900
G W S L R E T A C L G K S Y A Q M W S L M Y F H R R D L K L A A N	3268
GCUAUUUGCUGCGGAGUACCAUUGGAGG	9999
A I C S A V P S H W V P T S R T T W S I H A K H E W H T T E D M L	3301
ACAGUCUGGAAAGGUGGAAUUAAGAAACCAUUGGAGG	10098
T V W N R V W I Q E N P W M E D K T P V E S W E E I P Y L G K R E	3334
GACCAAUGGCGGCGUUAUUGGAGG	10197
D Q W C G S L I G L T S R A T W A K N I Q A A I N Q V R S L I G N	3367
GAAGAAUACAGAUUAUGGAAAGAAUUGGAAAGCAUACAACAAGGAAACCUUACAGAGGAAUUGGAAUUGGAAUUGGAAUUGGAAUUGGAAUUGGAAUUGG	10296
E E Y T D Y M P S M K R F R R E E E E A G V L W *	3391
GCUAGAAUCAGGUGGAAUCCAGCAUAGUACGGAAGAAACUAGGUCUCCGUGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	10395
CAUAGCUUAGUAAAUAUGCAGGCGUAGCUCACCGGAGG	10494
ACUAGCGGUUAGAGGAGACCCUCCUUAACAAGG	10593
AGGAGACCCCGGAAACAAAACAGCAUUAUGCAGG	10692
AUGGAAUGGUGUGUUAUUAACAGGUUCU	10723

Fig. 2—Continued

TABLE 1

NUCLEOTIDE AND AMINO ACID DIFFERENCES BETWEEN THE PARENTAL DENGUE-2 16681 AND ITS 16681-PDK 53 VIRUSES

Gene	No.	Nucleotide		Amino acid	
		16681	16681-PDK53	16681	16681-PDK53
5' untrans	57	C	U		
C	115	A	G	K	E
	210	A	C		
	226	C	U		
	234	G	A		
PrM(-M)	524	A	U	D	V
	651	C	G		
	703	C	G	Q	E
M	903	C	A	F	L
E	1317	A	G		
	1445	C	U	T	I
	1477	G	A	V	I
	1509	C	A		
	1524	G	C	M	I
	1557	G	C		
	2055	C	U		
NS1	2475	G	U		
	2579	G	A	G	D
	2821	C	U	L	F
	3081	C	U		
	3129	G	A		
	3147	A	U		
	3178	C	A	Q	K
NS2A	3620	U	G	M	R
	3776	C	A	A	E
	3824	C	U	T	I
	3987	G	A		
NS2B	4241	C	U	P	L
	4344	C	U		
	4433	C	G	T	R
NS3	4824	G	C	E	D
	5649	C	G	S	R
	5705	C	U	A	V
	5955	C	U		
NS4A	6442	G	A	D	N
	6580	C	U	L	F
	6599	G	C	G	A
NS4B	7403	U	C	V	A
	7439	G	C	W	S
	7561	G	A	A	T
NS5	7571	G	U	G	V
	7578	C	U		
	8571	C	U		
	9221	C	A	A	E
	9699	U	A		
	9732	C	U		
	9845	A	C	D	A
3' untrans	10316	U	C		
	10330	C	G		
	10637	C	A		
	10649	C	A		
	10651	A	U		
	10673	U	C		

TABLE 2

SUMMARY OF THE GENETIC DIFFERENCES BETWEEN THE DENGUE-2 16681 AND 16681-PDK 53 VIRUSES

Gene/protein	Nucleotide differences (%)		Amino acid differences (%)	
5' untranslated	1/96	(1.04)	—	
C	4/342	(1.17)	1/114	(0.88)
prM	3/273	(1.10)	2/91	(2.20)
M	1/225	(0.44)	1/75	(1.33)
E	7/1485	(0.47)	3/495	(0.61)
NS1	7/1056	(0.66)	3/352	(0.85)
NS2A	4/654	(0.61)	3/218	(1.38)
NS2B	3/390	(0.77)	2/130	(1.54)
NS3	4/1854	(0.22)	3/618	(0.49)
NS4A	3/450	(0.67)	3/150	(2.00)
NS4B	3/744	(0.40)	3/248	(1.21)
NS5	7/2700	(0.26)	3/900	(0.33)
3' untranslated	6/454	(1.32)	—	
Total	53/10723 (0.49)		27/3391 (0.80)	

100 units/ml penicillin, and 100 µg/ml streptomycin. At 70% confluency, the cells were infected with the dengue-2 16681 or 16681-PDK 53 virus in RPMI 1640 containing 0.2% bovine serum albumin (BSA). Hemagglutinin (HA) was assayed as described by Clarke and Cassals (1958) and tissue culture fluids containing the virus particles were collected for 3 consecutive days, at peak HA titers. After centrifugation to remove cells and debris, the virions were precipitated by adding  $\frac{1}{5}$  volume of 40% polyethylene glycol, 15% NaCl, and incubated at 4° for 1 hr. This precipitate was then collected by centrifuging at 7500 g for 30 min at 4° and resuspending in buffer (0.1 M NaCl, 0.012 M Tris-HCl, pH 7.5, 0.2% BSA) before centrifuging in a 5–50% sucrose gradient. The visible band of virions was collected and viral RNA extracted in 0.05 M Na acetate, 0.1 mM EDTA solution containing 0.5% SDS, and an equal volume of phenol/chloroform. The RNA was recovered from the aqueous phase by ethanol precipitation. It was ethanol precipitated twice before use for cDNA synthesis.

### Synthesis of cDNA

cDNA was synthesized as described previously (Blok *et al.*, 1984). Viral RNA and either a 14-mer based on the West Nile virus sequence (Wengler and Wengler, 1981) or oligo (dT)<sub>12–18</sub> were heated to 90° for 1 min, quickly chilled in an ice/water bath, and used as a template for cDNA synthesis. For synthesis, [<sup>32</sup>P]-dATP (Bresatec) and avian myoblastosis virus reverse transcriptase (AMV RT, Life Sciences) were added and

incubated at 37° for 1 hr. The cDNA-RNA hybrids produced by this reaction were then used for synthesis of double-stranded DNA (dsDNA) using RNase H and DNA polymerase (Boehringer-Mannheim) as described by Okayama and Berg (1982). The resulting dsDNA was cleaved with the restriction enzymes *Sau3A* or *Bam*HI, inserted into the *Bam*HI site of the plasmid vector pUC8, and cloned using standard methods (Maniatis *et al.*, 1982).

#### Amplification of DNA using the polymerase chain reaction (PCR)

Oligonucleotide primers (20–23-mers with an average GC content of 60%) flanking the regions to be amplified were made in a Milligen oligonucleotide synthesizer (University of Queensland), desalted in a PD-10 Sephadex G-25 *M* column (Pharmacia), and the first three fractions were dried and resuspended in 100  $\mu$ l water. A reaction mixture containing 1  $\mu$ l (ca. 100 pmol) of each of the two primers, 400 ng viral RNA in 1.5 mM MgCl<sub>2</sub>, 50 mM KCl, 10 mM Tris-HCl, pH8.0, and 0.02% BSA was heated to 95° for 1 min and cooled to 42° over 3 min. AMV RT (2 units) and *Taq* polymerase (2 units, Cetus) were added to the mixture and this was incubated at 42° for 5 min. The DNA was then amplified for 30 cycles (95° for 1 min, 55° for 2 min, and 72° for 3 min). The resulting fragments were purified on a polyacrylamide gel, eluted in buffer (0.1 *M* Tris-HCl, pH8.0, 0.5 *M* NaCl, 5 mM EDTA) for 16 hr at room temperature, phenol extracted, and ethanol precipitated prior to sequencing.

#### Sequencing of cloned DNA and PCR fragments

Sequences were obtained by the dideoxy chain termination method as described (Sanger *et al.*, 1977). In a typical experiment, 4  $\mu$ g of circular plasmid DNA or purified PCR fragment DNA was added to 40  $\mu$ l of 0.2 *M* NaOH, 0.2 mM EDTA in order to separate the strands. This mixture was incubated at room temperature for 5 min, precipitated with 3 vol of ethanol in the presence of 37.5 mM Na acetate pH5.5, and then collected by centrifugation in a microfuge for 5 min. The resulting pellet was dried in a vacuum and resuspended in 20  $\mu$ l sterile water. This treated DNA was enough for five sequence experiments and could be stored up to 2 weeks at -20°. Treated DNA (4  $\mu$ l or 0.8  $\mu$ g) was mixed with 1  $\mu$ l (1 pmol) of primer, heated to 95° for 1 min, quickly chilled in an ice/water bath, and used in the dideoxy sequencing procedure in the presence of [<sup>32</sup>P]dATP (Bresatec). In most cases 0.5 unit DNA polymerase (Klenow fragment, Boehringer-Mannheim) was used to obtain sequence data, but for some of the larger clones, 3 units of Sequenase (Trace) was

used. The reactions were stopped by the addition of formamide containing bromophenol blue and xylene cyanol dye, heated to 95° for 3 min, and chilled prior to electrophoresis in 8% polyacrylamide gels containing 7 *M* urea. The gels were fixed in 10% acetic acid, wrapped in Saran Wrap, and autoradiographed at room temperature without intensifying screens.

#### Viral RNA sequencing

Dengue-2 genomic RNA was sequenced using the dideoxy chain termination method (Sanger *et al.*, 1977) as described (Air, 1979). Briefly, oligonucleotide primers, based on sequence data obtained from cDNA clones or published dengue-2 sequences, were synthesized on a Biosystems oligonucleotide synthesizer (Qld Inst Med Res), desalted in a PD-10 Sephadex G-25 *M* column (Pharmacia), and the first three fractions were dried and dissolved in 100  $\mu$ l water. Primer (usually 1  $\mu$ l of 1:100 dilution, although the dilution was determined empirically for each primer) and 0.05  $\mu$ g RNA were heated to 90° for 1 min, quickly chilled in ice/water, and sequenced with 1.0 unit of AMV RT (Life Sciences).

#### Computing

Sequences were aligned, compared, and classified using various programs of the RSBS SEQ, the RSBS STATS, and the GCG (Version 6.2; Devereux *et al.*, 1984) libraries. The progressive alignment program (Feng and Doolittle, 1987, 1990) with the MDM78 similarity matrix (Dayhoff *et al.*, 1978) and the neighbor-joining tree program (Saitou and Nei, 1987; Studier and Keppler, 1988) were also used. The FASTA program of the GCG library was used to search Release 16 (November 1990) of the SwissProt database for proteins with sequences related to those of flaviviruses.

## RESULTS

#### Nucleotide and encoded amino acid sequences of dengue-2 16681 and 16681-PDK53 viruses

The sequencing strategies used to obtain the complete nucleotide sequences of the dengue-2 16681 and 16681-PDK53 viruses are shown in Fig. 1. This figure shows that the sequences of the E and NS1 genes were obtained from both cloned material and from sequencing the RNA genome directly. Other regions which were not isolated during the preparation of the cDNA libraries were also sequenced from the viral RNA using specific oligonucleotide primers based on sequences from the dengue-2 16681, 16681-PDK53 or Jamaica (Deubel *et al.*, 1986, 1988) strains. The cDNA clones were obtained using either oligo dT<sub>12-18</sub>



or the synthetic oligonucleotide 5' AGAACCTGTTGATT 3' as a primer for cDNA synthesis followed by *Sau3A* or *Bam*HI digestion. Both the 16681 and 16681-PDK53 cDNA libraries contained multiple copies of most of the cloned fragments, although there was a unique *Bam*HI clone in the 16681-PDK53 library. The cDNA clones were sequenced from both strands and represented 79 and 74% of the 16681 and 16681-PDK53 RNA genome, respectively.

The 3' terminal primer, 5' AGAACCTGTTGATT 3', was used to sequence the extreme 3' end of the viral RNA but the sequence of only 79 nucleotides could be determined. The rest of the 3' nontranslated region proved to be difficult to sequence with specific primers based on the dengue-2 Jamaica sequence (Deubel *et al.*, 1988), probably because of sequence mismatching (when the sequences of Jamaica and 16681/16681-PDK53 were compared, 19/15 of the 454 nucleotides differed). This region at the 3' end of the 16681-PDK53 genome was therefore amplified using the polymerase chain reaction (PCR) and the primers 5' CTTTCCCAGCGTCAATATGCTG 3' (based on sequence of the viral RNA) and 5' GGCTAGAAGTCAGGTCCGATC 3' (based on sequence of a 16681-PDK53 cDNA clone). The resulting 344 bp fragment (nucleotides 10296 to 10649, see Fig. 2) was produced in five different PCR experiments, and each of these was sequenced to ensure that any possible errors produced during amplification by the *Taq* polymerase could be detected. However, the five sequences were identical. For the 16681 virus, the sequence of this 3' noncoding region and the 3' portion of the NS5 gene were obtained by amplifying the same 344 bp fragment using PCR as well as a 550 bp fragment using the primers 5' CCTTGACGGGGCTCACAGGTAG 3' (based on sequence of 344 bp PCR fragment) and 5' GGGAGACGGCCTGTTTGGGG 3' (based on sequence 16681-PDK53 cDNA clone). The resulting sequence from these two amplification reactions encoded the region from nucleotides 9815 to 10649.

The entire nucleotide sequence of the dengue-2 16681-PDK53 viral RNA genome and its encoded amino acid sequence are shown in Fig. 2 and it represents the consensus RNA sequence obtained using three different strategies (Fig. 1): (i) cDNA cloning, (ii) sequencing particular regions of the viral RNA genome directly, and (iii) amplification of a fragment at the 3' end using PCR. The 3' terminal 14 nucleotides, complementary to the synthetic primer based on the West Nile virus sequence, have not been deduced but have simply been added to the 3' terminal deduced nucleotide since this primer hybridized very well both dengue-2 16681 and 16681-PDK53 RNA. The differences between the 16681-PDK53 sequence (Fig. 2) and the

sequence of its parental virus 16681 are shown in Table 1.

The N-terminal amino acid sequences of the prM, M, and E proteins of the 16681 virus were determined using a gas phase protein sequencer coupled to reverse phase ODS columns. These confirmed the encoded sequences as well as those reported previously for dengue-2 proteins.

### Comparison of the 16681 and 16681-PDK53 nucleotide sequences

There were 53/10723 (0.49%) nucleotide differences between the parental dengue-2 16681 and its candidate vaccine derivative 16681-PDK53 virus, of which about half (26/53) were silent. Table 2 highlights the differences in the genes and proteins of these two viruses and together with Table 1 show that the differences occur throughout the entire genome in both the coding and noncoding regions. Nucleotide changes occurred more frequently in regions such as the 5' and 3' untranslated regions (1.04% and 1.32% or 1/96 and 6/454 nucleotide differences, respectively), the C (1.17% or 4/342 nucleotide differences) and the prM portion of the prM/M gene (1.10% or 3/273 nucleotide differences). A disproportionate number of nucleotide differences which led to amino acid substitutions (14/17 or 88%; which can also be expressed as 12% silent mutations) was detected in the prM/M, NS2, and NS4 regions. The other proteins had a higher percentage silent mutations (55%).

### Differences in the amino acid sequence

Tables 1 and 2 also summarize the amino acid differences between the virulent and attenuated dengue-2 viruses. Of the virion proteins, the prM peptide of the prM/M protein has changed to the greatest extent (2.20% or 2/91 amino acid differences). One of the amino acid substitutions (Asp to Val, 16681 to 16681-PDK53, respectively) changed the character of the residue from hydrophilic to hydrophobic and could therefore be important. Only one amino acid substitution (Phe to Leu, a conservative change) was present in the M peptide of the prM/M protein (1.33% or 1/75 amino acid differences). A 0.88% amino acid difference (1/114 amino acids) with a change of charge from a basic residue to an acidic one (Lys to Asp) was found in the capsid protein. The other virion protein, the envelope glycoprotein, had 3/495 amino acid substitutions (or 0.6% difference).

NS1 had 7/1056 nucleotide changes and 3/352 amino acid changes. One of the three amino acid substitutions changed a small residue, Gly, to an acidic residue, Asp. The non-structural proteins, NS2 and

TABLE 3  
CODES AND REFERENCES OF THE FLAVIVIRUSES USED IN THIS STUDY

Virus isolate	Code name	References
Dengue-1—Nauru	D1	Mason <i>et al.</i> (1987)
Dengue-2—16681-PDK53	D2PDK	This work
Dengue-2—Jamaica 1409	D2JAM	Deubel <i>et al.</i> (1986, 1988)
Dengue-2—NGC	D2NGC	Irie <i>et al.</i> (1989)
Dengue-2—PR159S1	D2S1	Hahn <i>et al.</i> (1988)
Dengue-3—H87	D3	Osatomi and Sumiyoshi (1990)
Dengue-4—814669	D4	Mackow <i>et al.</i> (1987)
Japanese encephalitis—Nakayama	JE	Zhao <i>et al.</i> (1986)
Japanese encephalitis—JaOArS982	JEV	McAda <i>et al.</i> (1987)
Kunjin—MRM61C	KUN	Sumiyoshi <i>et al.</i> (1987)
Murray Valley encephalitis	MVE	Coia <i>et al.</i> (1988)
St. Louis encephalitis—MS1-7	SLE	Dalgarno <i>et al.</i> (1986)
Tick-borne encephalitis—ZZ-9	TBE	Trent <i>et al.</i> (1987)
Tick-borne encephalitis—USSR	TBEV	Mandl <i>et al.</i> (1988, 1989b)
West Nile	WNV	Yamshchikov and Pletnev (1988)
Yellow fever—17D vaccine	YF	Castle <i>et al.</i> (1985, 1986)
Yellow fever—Asibi	ASIBI	Wengler <i>et al.</i> (1985)
		Rice <i>et al.</i> (1985)
		Hahn <i>et al.</i> (1987)

NS4, contained a high proportion of amino acid differences (1.38 and 1.54% or 3/218 and 2/130 amino acids for NS2A and B; and 2.00 and 1.21% or 3/150 and 3/248 amino acids for NS4A and B, respectively). The NS3 and NS5 proteins were more highly conserved, showing a 0.49% (3/618 amino acids) and a 0.33% (3/900 amino acids) protein difference, respectively. The preservation of sequence was also observed at the gene level with a 0.22% (4/1854 nucleotides) and 0.26% (7/2700 nucleotides) difference for the NS3 and NS5 genes, respectively.

#### Relationship of the 16681/16681-PDK53 virus pair with other flaviviruses

The relatedness of dengue-2 16681/16681-PDK53 virus pair to other flaviviruses, and the relationships within the group, were determined using the protein sequences deduced from their published nucleotide sequences to make various comparisons. As the 16681 and 16681-PDK53 viruses were very closely related, only the 16681-PDK53 sequence (Fig. 2) was used in further analyses. The viral protein sequences compared were those of D1, D2JAM, D2NGC, D2PDK, D2S1, D3, D4, JEV, JE, KUN, MVE, SLE, TBE, TBEV, WNV, ASIBI and YF viruses (see Table 3 for full virus names and references). At the time of our analyses, only parts of the genomes of D1, MVE, and SLE were known to us, and so only some of their proteins were used in the comparisons.

The 11 flavivirus proteins (C, prM, M, E, NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) were analyzed sep-

arately. In each of the 11 sets the proteins were aligned, and the Feng, Johnson and Doolittle (1985) distance measures (FJD distances) between all pairs of each set were estimated after progressive alignment (Feng and Doolittle, 1987, 1990). The smallest FJD distances were between the NS5 proteins, and therefore the FJD distances for the others were, for each virus-virus comparison, converted into the distances relative to those of the NS5 comparisons. These ratios were averaged for each protein (Fig. 3) and showed that the amount of change of the proteins, relative to NS5, varied over almost a fourfold range. The E, NS1, and NS3 proteins differed only up to 1.60 times as much as

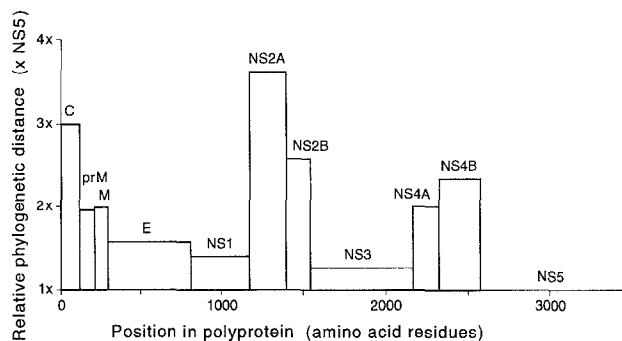
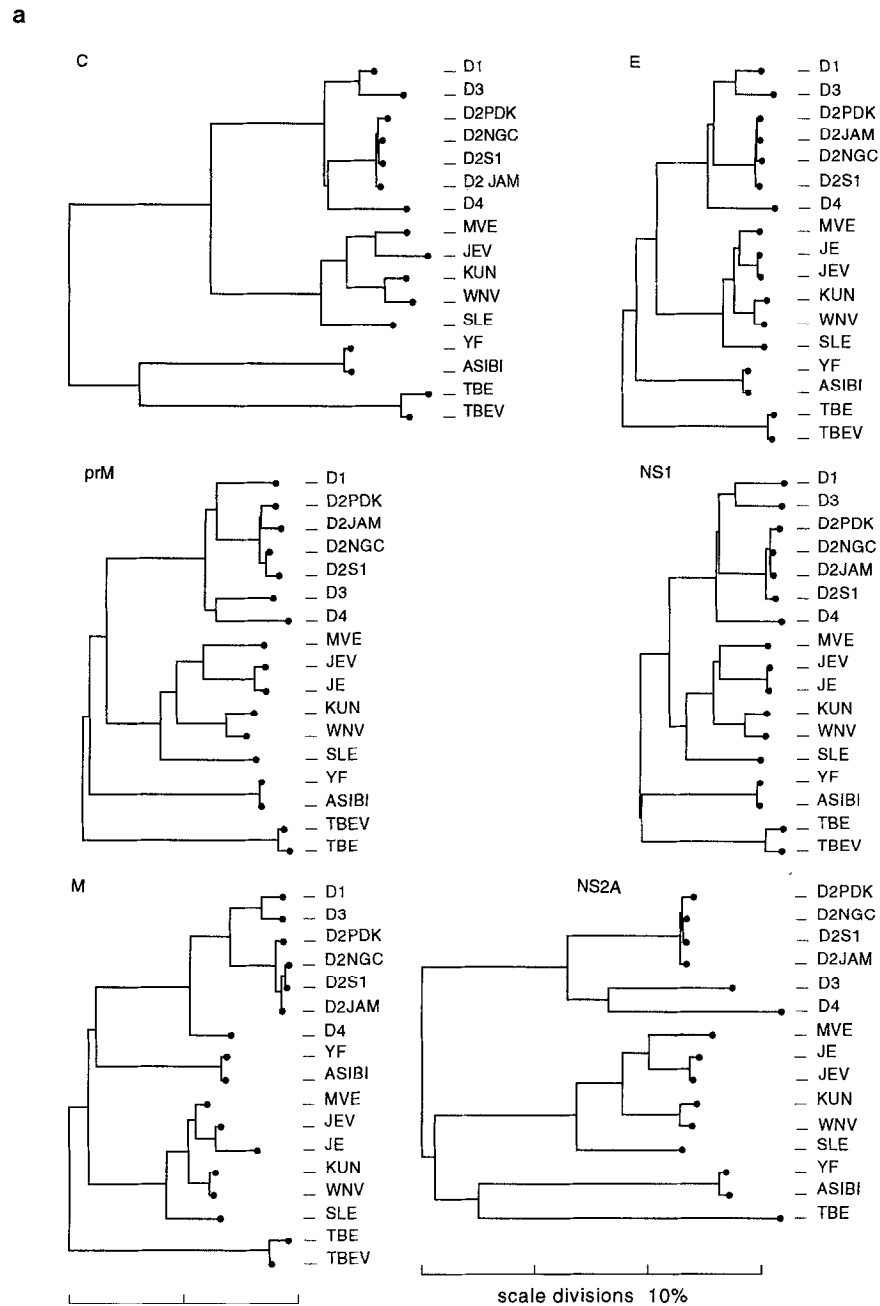


FIG. 3. Histogram showing the mean phylogenetic Feng, Johnson, and Doolittle (1985) distance measures for all pairwise comparisons of each protein from the 11 flaviviruses (dengue serotypes 1 to 4, Japanese encephalitis, Kunjin, Murray Valley encephalitis, St. Louis encephalitis, tick-borne encephalitis, West Nile, and yellow fever viruses) expressed in terms of the distances for NS5 comparisons.



**FIG. 4.** Dendrograms illustrating, for each flavivirus protein, the relationships of the amino acid sequences encoded for that protein in different flavivirus genomes (See Table 3 for full names and references of viruses). The sequences were aligned by the progressive-alignment method, and the phylogenetic Feng, Johnson, and Doolittle (1985) distance measures (MDM78 distance matrix) classified by the neighbor-joining method (Saitou and Nei, 1987). Scale units = 10%.

NS5, but the C, NS2A, NS2B, NS4A, and NS4B proteins differed by up to  $3.66 \pm 0.11$  times as much. This result is similar to that obtained by Miller and Purcell (1990), who calculated a histogram (their Fig. 3) of the invariant residues in the aligned proteins of eight flaviviruses.

Next, for each of the 11 proteins, a dendrogram representing the relationships of the individual proteins

was calculated by the neighbor-joining tree method (Saitou and Nei, 1987). The 11 dendrograms (Fig. 4) are closely similar in topology, especially those of the most slowly evolving proteins (i.e., E, NS1, NS3 and NS5), but their total branch lengths differ, thereby confirming the relative differences shown in Fig. 3. The dendrograms only differ in minor ways, mostly in their fine "twigs" (e.g., the relative positions of the dengue-

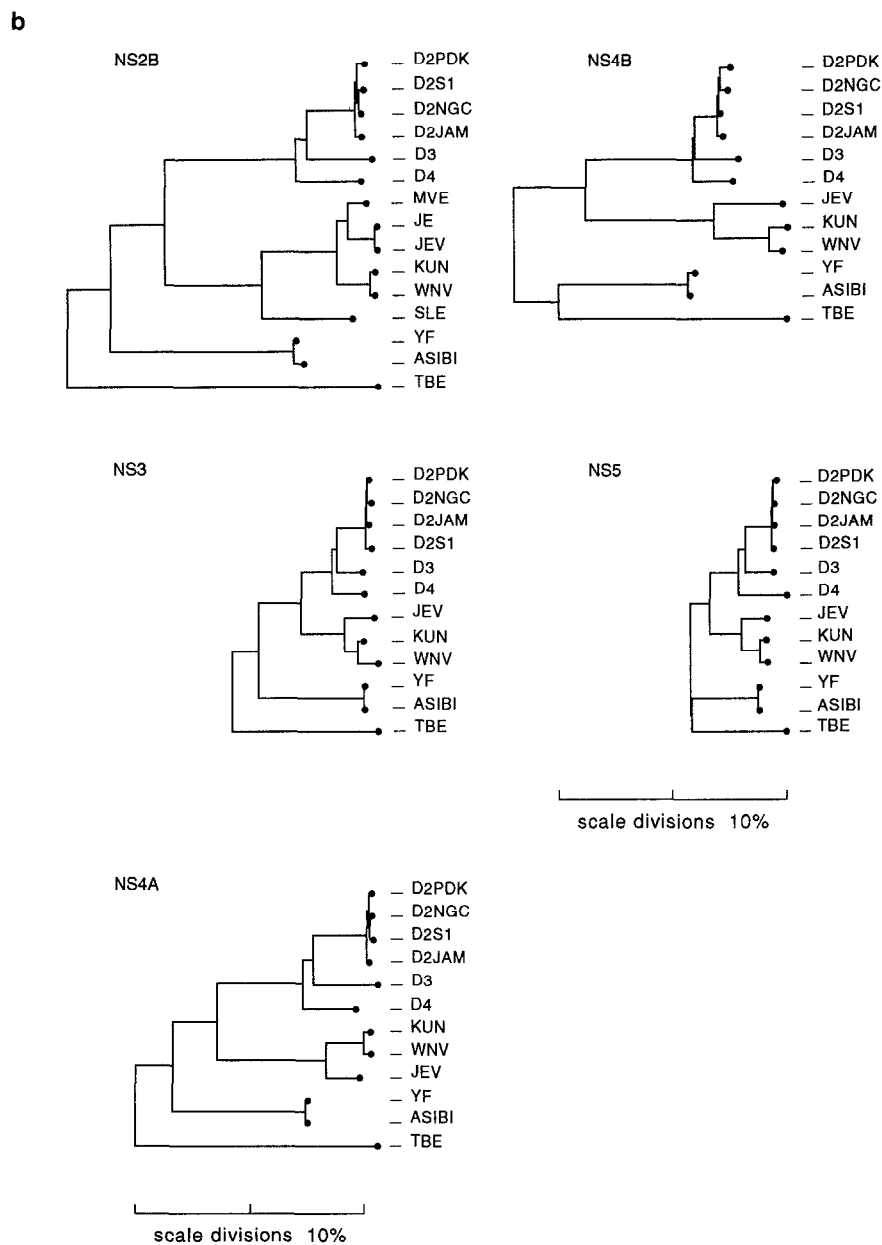


FIG. 4—Continued

2 isolates), and also in the position of the basal branch of each dendrogram, which was arbitrarily fixed as the point in the dendrogram that bisected the path linking the two most distant viruses.

As the analyses indicated that the flaviviruses had evolved by divergent mutational change, a single simple dendrogram was constructed (Fig. 5) by combining the distance matrices of the 11 proteins after correcting for their differences in size and relative rates of change. This dendrogram places the viruses into four clusters; the two TBE isolates, the two YF isolates, the six encephalitis viruses, and the seven dengue virus

isolates. This classification is similar to that reported by Mandl *et al.* (1989b), who used modified percentage identities of aligned sequences of the NS1 proteins of 11 flaviviruses to calculate a dendrogram by an undisclosed method. Our dendrograms, obtained by the neighbor-joining tree method, have given a more detailed binary branching pattern than that of Mandl *et al.* (1989b), which included two unresolved trifurcations.

#### Distant relationships

We examined the sequence similarities of the conserved flavivirus proteins (NS5 and NS3), with those of

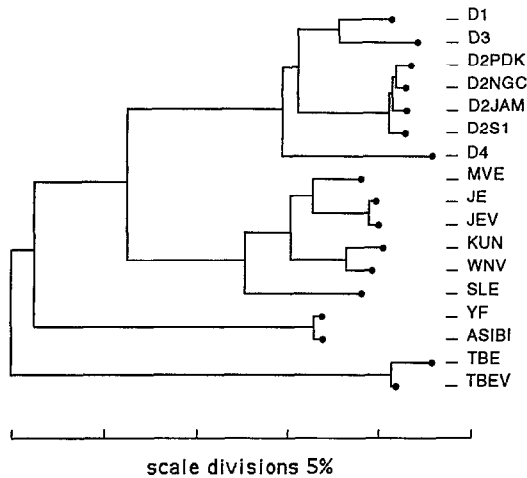


FIG. 5. Combined dendrogram illustrating the relationships of 17 flaviviruses (whose full names are shown in Table 3). Scale units = 5%.

other viruses and other organisms, to find the possible relatives of flavivirus genes. We also used the related proteins as "outgroups" to establish the position of the "root," and hence perhaps the "oldest" branching point, of the combined dendrogram of flaviviruses (Fig. 5).

The motif which includes the characteristic sequence -GDD- (or a close derivative of it) is found in all viral replicases (Argos, 1988; Bruenn, 1991; Diamond *et al.*, 1989; Gorbalenya and Koonin, 1988; Habili and Symons, 1989; Icho and Wickner, 1989; Kamer and Argos, 1984; Koonin *et al.*, 1989; Pietras *et al.*, 1988; Poch *et al.*, 1989), including the NS5 protein of flaviviruses.

We classified the -GDD- regions of 63 viral genomes including those of 12 flaviviruses. Each -GDD- region compared consisted of 300 amino acid residues, stretching from 240 residues on the N-terminal side of the -GDD- to 60 residues to its C-terminal side. These sequences were aligned pairwise using the Needleman-Wunsch algorithm (MDM78 matrix and gap penalty 6, Needleman and Wunsch, 1970) and the number of identical residues was recorded, as well as the difference between the alignment score and the mean score of 100 randomizations of the sequences expressed in standard deviation units (StDUs) calculated from the randomizations. Both sets of data were converted to distance matrices; the identical residue scores were converted into the percentages of non-identical residues ("% non-ident"), including gaps as a "21st" amino acid, and the standard deviation measures were converted into "standard deviation distances" (StDDs) between each pair of sequences using the formula  $StDD = 2.0 - \log_{10}(StDU + 3.0)$ . The

neighbor-joining tree method was then used to compute dendrograms from the distance measures.

Both dendrograms placed the "root" of the flavivirus tree on the link joining the TBE branch to that of the other flaviviruses. Figure 6 shows the % non-ident dendrogram. It placed the viruses into seven major groups. The flaviviruses were a sister group to the pestiviruses which then joined sequentially to the carmo-luteo(BYD) cluster and hepatitis C. These viruses then had the sobemo-luteo(BWY) cluster and black beetle nodavirus as a sister group, before being linked to the other viruses. The StDD dendrogram placed the viruses (except black beetle nodavirus and Q $\beta$  levivirus) into the same seven major groupings as the % non-ident dendrogram, but the groupings were linked differently so

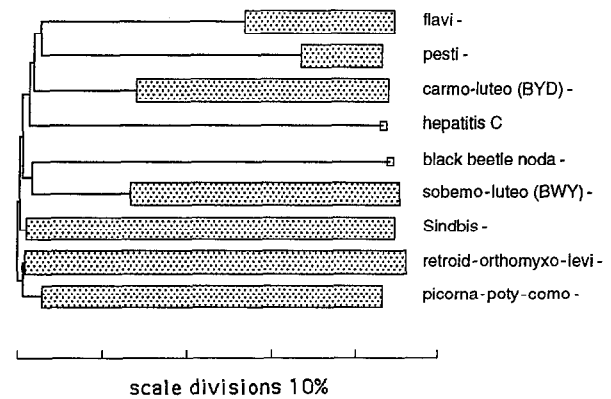


FIG. 6. Dendrogram showing the relationships of the RNA polymerase motifs of the genomes of 63 viruses. Each shaded rectangle represents a group of viruses, and illustrates the variability of the group; the left edge of the rectangle is placed where the final node fusing members of the group occurs, and the right edge is the position of the individual on the longest branch. Members of the groups are: *flavi-*: ASIBI, D2JAM, D2NGC, D2PDK, D2S1, D3, D4, JEV, KUN, TBEV, WNV, YF (see Table 3 for full names of viruses); *pesti-*: bovine viral diarrhea and hog cholera pestiviruses; *carmo-luteo(BYD)-*: barley yellow dwarf (PAV) and soybean dwarf luteoviruses, red clover necrotic mosaic dianthovirus, carnation mottle, melon necrotic spot and turnip crinkle carmoviruses, maize chlorotic mottle virus, cucumber necrosis, cymbidium ringspot and tomato bushy stunt tobamoviruses, tobacco necrosis necrovirus; *sobemo-luteo(BWY)-*: beet western yellows, potato leaf roll (two isolates) luteoviruses, southern bean mosaic sobemovirus (two isolates); *Sindbis-*: alfalfa mosaic virus, brome mosaic bromovirus, cucumber mosaic cucumovirus, barley stripe mosaic hordeivirus, tobacco rattle and pea early browning tobamoviruses, tobacco mosaic tobamovirus, Sindbis alphavirus, eggplant mosaic, kennedy yellow mosaic, ononis yellow mosaic and turnip yellow mosaic tymoviruses, potato X and white clover mosaic potexviruses, potato M and potato S carlaviruses, beet necrotic yellow vein furovirus; *reteroid-orthomyxo-levi-*: human immunodeficiency lentivirus, influenza A orthomyxovirus, mouse hepatitis coronavirus, cauliflower mosaic caulimovirus, Q $\beta$  levivirus; *picorna-poty-corno-*: rhinovirus 2, poliovirus, encephalomyocarditis cardiovirus, foot-and-mouth disease aphthovirus, cowpea mosaic comovirus, plum pox, potato Y, tobacco etch and tobacco vein mottling potyvirus. Scale units = 10%.

that although the pestiviruses and hepatitis C were sister groups, they were further from the flaviviruses. The reasons for the differences between these dendrograms are not known, but it could have resulted from the decision to use a segment of fixed length from each sequence, rather than attempting to align subjectively some very tenuously homologous sequences.

The nucleotide binding and helicase motifs, which have the core sequence -GKT/S-, are found in many proteins including those of viruses (Walker *et al.*, 1982; Higgins *et al.*, 1986; Gorbalenya *et al.*, 1989). This motif is found in the C-terminal part of the NS3 protein of flaviviruses, and the region has been shown (Gorbalenya *et al.*, 1989; Miller and Purcell, 1990) to have sequence similarities to the NS3-like protein of hog cholera pestivirus and hepatitis virus C and, more distantly, to the cylindrical inclusion (CI) proteins of the potyviruses (Bazan and Fletterick, 1989). Recently, Lain *et al.* (1989) have confirmed that the potyviral CI proteins have helicase activity, and Chambers *et al.* (1990) have shown that the NS3 protein of YF is a viral protease for specific post-translation processing of the viral polyprotein. A search of the Swissprot database (Release 16) by the FASTA technique using the NS3 proteins of D2PDK and YF failed to identify any other proteins with sequences of more than 100 amino acid residues with significant similarity; several much shorter sequences with clear similarity were found.

Therefore a classification was computed of five flavivirus NS3 proteins (D2PDK, D3, D4, YF, and TBE), the NS3-like proteins of the two pestiviruses and hepatitis C virus, and the CI proteins of plum pox (strain D; Teycheney *et al.*, 1989), potato virus Y (strain N; Robaglia *et al.*, 1989), tobacco etch (Allison *et al.*, 1986), and tobacco vein mottling (Domier *et al.*, 1986) viruses using the progressive-alignment/FJD distance/neighbor-joining tree method. The classification further confirmed that the flavivirus tree is rooted near the base of the TBE branch. The classification showed clearly that the potyvirus CI proteins are only very distantly related to the NS3 and NS3-like proteins. The dendrogram was, in essence, H-shaped with a short central link of 20% with branches at one end to the pestivirus and hepatitis C proteins of 78 and 98% respectively, and, from the other end, of 31 and 155% to the flavivirus and potyvirus proteins.

Searches of the Swissprot database (Release 16) and the NS1 and E proteins of D2PDK and YF using the FASTA technique found only a few short sequences (<100 residues) with similarity to the NS1 proteins, but a clear sequence similarity between the flavivirus E proteins and the spike glycoprotein precursors of vesicular stomatitis rhabdovirus, strains Orsay and Glasgow (Gallione and Rose 1985; Vandepol and Holland, 1986)

was found. The alignment scores of the E and spike proteins by the Needleman-Wunsch (1970) method were 2.3–2.9 standard deviations from the mean alignments scores of 100 randomizations of the same sequences.

## DISCUSSION

Differences in the nucleotide and encoded amino acid sequences occurred throughout the genomes of the virulent (16681) and attenuated (16681-PDK53) dengue-2 virus strains. Certain genes (M, NS4A, and NS4B) displayed nucleotide changes which all led to amino acid substitutions, while others (C, E, NS1, and NS5) underwent a majority of silent nucleotide changes. However, it is not possible to predict from these changes which particular protein(s) and/or region(s) of the genome are involved in the reported attenuation of 16681-PDK53.

Monoclonal antibodies to at least three dengue virus proteins, namely, prM, E, and NS1, have been shown to protect mice against challenge with dengue viruses (Brandiss *et al.*, 1986; Henchal *et al.*, 1988; Kaufman *et al.*, 1987, 1989; Schlesinger *et al.*, 1987; Zhang *et al.*, 1988). There are nucleotide and amino acid differences between the 16681 and 16681-PDK53 viruses in these proteins but their significance is not obvious.

The prM-specific monoclonal antibodies which protected mice against challenge with both homologous and heterologous dengue viruses (Kaufman *et al.*, 1989) did not display neutralizing activity *in vitro*. The locations of the protective epitopes within the prM peptide have not yet been identified and it is therefore not known whether the amino acid differences (Table 1) found between the 16681 and 16681-PDK53 viruses in this protein are related to protection.

The E protein is the major dengue virus antigen and antibodies to this protein neutralize the virus. Proposed three-dimensional structures for the E protein of related flaviviruses, WN virus (Nowak and Wengler, 1987) and TBE virus (Mandl *et al.*, 1989a) have been developed based on monoclonal antibody-binding and proteolytic cleavage studies. Adapting the dengue-2 envelope protein sequence to the flavivirus models shows that the three amino acid changes are in stem regions. It seems unlikely that the differences in the E protein are significant as two of the changes (Val to Ile and Met to Ile) are conservative changes which occur in the stem structures of domain A, rather than in the more accessible loops (based on the model of the E protein of TBE virus, Mandl *et al.*, 1989a). Interestingly, the other amino acid substitution (Thr to Ile) is in the same position as an antigenic variant reported for TBE virus (Lys to Glu) (Mandl *et al.*, 1989a); and this Thr to

Ile substitution occurs in a region of dengue-2 which reacts with sera from a range of dengue-2-infected hosts (Aaskov *et al.*, 1989). The positions of the three amino acid changes are also outside any proposed flavivirus group reactive or dengue-2 specific continuous epitopes proposed by Innis *et al.* (1989).

Twelve amino acid differences were detected between the E proteins of the virulent (ASIBI) and an attenuated 17D yellow fever viruses (Hahn *et al.*, 1987). This contrasts with five amino acid differences observed in the E protein of a virulent and an attenuated strain of JE virus (Nitayaphan *et al.*, 1990) and the three amino acid differences reported here for the dengue-2 16681 (virulent) and dengue-2 16681-PDK53 (attenuated) viruses. This difference in number of amino acid changes between the parent and attenuated strains (12 for YF pair, 5 for JE pair, and 3 for the dengue-2 pair) may simply reflect the different clinical syndromes and pathogenesis of these viruses, or the different passage histories of these viruses [234 passages for YF pair (Theiler and Smith, 1937), 122 passages for the JE pair (Eckels *et al.*, 1988) and 53 passages for the dengue-2 pair (Yoksan *et al.*, 1986)], or the fact that the YF and JE viruses were plaque purified before cDNA cloning and sequencing whereas the dengue-2 viruses used for cloning and sequencing represent the virus population as a whole.

Monoclonal antibodies to the NS1 protein protected mice against lethal challenge in experiments with YF virus (Gould *et al.*, 1986; Schlesinger *et al.*, 1985, 1986) and dengue-2 (Henchal *et al.*, 1988; Schlesinger *et al.*, 1987) indicating that NS1 may be important in flavivirus infections. There are three amino acid substitutions (Table 1) and one of these (Gly to Asp) may be significant because this Gly is present in all other flaviviruses sequenced to date, i.e., D1, D2JAM, D2NGC, D2S1, D3, D4, JEV, JE, KUN, MVE, SLE, TBE, TBEV, WNV, ASIBI, and YF viruses; see Table 3 for full virus names and references. However, the three-dimensional structure of NS1 is not known and, as no model has been proposed, it is difficult to assess the significance of this substitution.

The function of the NS2 and NS4 proteins in virus replication is unknown but their hydrophobic profiles are remarkably conserved among both the dengue viruses and flaviviruses in general, although their amino acid sequences are not. The amino acid substitutions (Table 1) in the NS4 region would have only marginal effects on this hydrophobicity profile but the NS2 regions involve substitutions of hydrophilic residues for hydrophobic ones, e.g., Met to Arg and Ala to Glu in NS2A. The significance of these results is difficult to determine since the function of these hydrophobic regions is unknown.

The NS3 and NS5 proteins are important in viral replication; and NS3 has been shown to contain sequences characteristic of a trypsin-like serine protease (Bazan and Fletterick, 1989) and is used in post-translational processing of the YF polyprotein (Chambers *et al.*, 1990). As these proteins possess important enzymatic functions, it can be expected that they can tolerate fewer amino acid substitutions. This is indeed what was observed with only 0.49 and 0.33% amino acid differences for NS3 and NS5, respectively, although there were substitutions which altered the charge (e.g., Ala to Glu or Asp to Ala) in NS5 protein.

Now that infectious virus from a cloned full-length cDNA copy of YF virus (Rice *et al.*, 1989) and a dengue-4 virus (Lai *et al.*, 1991) have been rescued from cells, it should be possible to do the same with the dengue-2 virus. *In vitro* mutagenesis can then be used to determine which of the above change or changes in the dengue virus genome and its encoded proteins relates to attenuation of the virus.

Analyses of the 11 dengue-2 16681-PDK53 encoded proteins (C, prM, M, E, NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) with the 16 published flavivirus sequences have shown that the dendrograms of the separate proteins (Fig. 4) are very similar in topology. The dendrograms of the most slowly evolving proteins (i.e., E, NS1, NS3, and NS5) give almost identical classifications although some branch points are in slightly different positions. There were however, no clear differences in the branching order of the different proteins which would indicate the occurrence of genetic recombination. It is therefore justifiable to calculate a combined dendrogram of these flaviviruses (Fig. 5) as it seems that the flaviviruses have evolved by divergent mutational change and there is no evidence of genetic recombination which has been shown to be a major factor in the evolution of RNA genome viruses; especially the retroviruses (McClure *et al.*, 1988), but also the luteo-like viruses whose replicase is so similar to that of the flaviviruses (see Fig. 6; Bruenn, 1991).

We believe that features of the combined dendrogram accurately reflect the flavivirus relationships, and hence the likely phylogeny, although there is at present no published method for testing the statistical significance of the details of neighbor-joining trees. These consistent features are that (i) the replicate isolates of individual viruses [i.e., three viruses (Japanese encephalitis, yellow fever, and tick-borne encephalitis) represented by two isolates each, and one (dengue-2) represented by four] classify together before they are linked in higher order relationships with other viruses; thus, the sequence data relationships are confirmed by the prior serological identification of individual isolates. (ii) The separate dendrograms of the 11 proteins (i.e., in-

dependent subsamples of the data) give closely similar dendrograms, indeed a dendrogram with the same topology could have been obtained by qualitatively drawing a consensus dendrogram from the eleven separate dendrograms.

The analyses of -GDD- motif of the polymerases of 63 viruses using different measures gave slightly different results and clearly show that, as expected, the results of analyzing distant sequence relationships must be interpreted with caution. Bruenn (1991) calculated a classification of the -GDD- motif regions which placed the flaviviruses together with hog cholera pestivirus as a sister group of the luteoviruses and relatives and, in order of increasing distance, black beetle nodavirus, then the tobamo-ARs, the alphaviruses, all the viruses with dsRNA genomes, and finally the picornaviruses and the leviruses. Our dendrogram (Fig. 6) shares some features with that of Bruenn (1991), but has the more agreed arrangements of the tymo-, tricorna-, poty-, and picornaviruses and confirms the report of Miller and Purcell (1990) of a shared sequence in the replicase proteins of hepatitis C virus and carnation mottle carmovirus (Guilley *et al.*, 1985).

Both dendrograms of the NS5 proteins, and that of the NS3 proteins, agreed in placing the root of the flavivirus tree close to the point where the TBE branch joined the other flaviviruses. Thus the "balance" of the unrooted combined tree (Fig. 5) is probably correct; it is rooted at its leftmost vertical link. This provides some evidence that the primary division of the flaviviruses was between those transmitted by mosquitoes and those transmitted by ticks, but this needs to be confirmed by the genomic sequencing of more of the tick-borne flaviviruses.

The distant relationships of the flavivirus proteins to proteins of other seemingly unrelated viruses add further to the evidence that most major virus groups have arisen by the recombination of viral genes from diverse sources. For instance: (i) the flavivirus E glycoprotein shows some sequence similarity to the spike glycoprotein of the vesicular stomatitis rhabdovirus; (ii) the flavivirus replicase (NS5) is close to that of the carmoviruses, yet their major virion proteins are quite different; that of the carmoviruses is an 8-stranded antiparallel  $\beta$ -barrel protein (a gene family that has provided the virion proteins of most small isometric virions); and (iii) the protease-like NS3 protein of flaviviruses shares a common ancestor with a protein of the picornaviruses and potyviruses that have unrelated coat proteins (Rossman and Rueckert, 1987; Dolja *et al.*, 1991).

#### ACKNOWLEDGMENTS

The authors thank Dr. Barry Gorman for useful comments. This research was supported by the World Health Organization as part of

its Program of Vaccine Development, and by the National Health and Medical Research Council of Australia, the Rockefeller Foundation, and the Sir Albert Sakzewski Foundation.

#### REFERENCES

- AASKOV, J. G., GEYSEN, H. M., and MASON, T. J. (1989). Serologically defined linear epitopes in the envelope protein of dengue 2 (Jamaica strain 1409). *Arch. Virol.* **105**, 209–221.
- AIR, G. M. (1979). Nucleotide sequence coding for the "signal peptide" and the N-terminus of the hemagglutinin from an Asian (H2N2) strain of influenza virus. *Virology* **97**, 468–472.
- ALLISON, R., JOHNSTON, R. E., and DOUGHERTY, W. G. (1986). The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: Evidence for the synthesis of a single polypeptide. *Virology* **154**, 9–20.
- ARGOS, P. (1988). A sequence motif in many polymerases. *Nucleic Acids Res.* **16**, 9909–9916.
- BAZAN, J. F., and FLETTERICK, R. J. (1989). Detection of a trypsin-like serine protease domain in flaviviruses and pestiviruses. *Virology* **171**, 637–639.
- BHAMARAPRAVATI, N., YOKSAN, S., CHAYANIYAYOTHIN, T., ANGSUBPHAKORN, S., and BUNYARATVEJ, A. (1987). Immunization with a live-attenuated dengue-2 virus candidate vaccine (16681-PDK53): Clinical, immunological and biological responses in adult volunteers. *Bull. Wld. Hlth. Org.* **65**, 189–195.
- BLOK, J., HENCHAL, E. A., and GORMAN, B. M. (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.
- BRANDISS, M. W., SCHLESINGER, J. J., WALSH, E. E., and BRISSELLI, M. (1986). Lethal 17D yellow fever encephalitis in mice. I. Passive protection by monoclonal antibodies to the envelope proteins of 17D yellow fever and dengue 2 viruses. *J. Gen. Virol.* **67**, 229–234.
- BRUENN, J. A. (1991). Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases. *Nucleic Acids Res.* **19**, 217–226.
- CASTLE, E., NOWAK, T., LEIDNER, U., WENGLER, G., and WENGLER, G. (1985). Sequence analysis of the viral core protein and the membrane-associated proteins V1 and NV2 of the flavivirus West Nile virus and of the genome sequence for these proteins. *Virology* **145**, 227–236.
- CASTLE, E., LEIDNER, U., NOWAK, T., WENGLER, G., and WENGLER, G. (1986). Primary structure of the West Nile flavivirus genome region coding for all nonstructural proteins. *Virology* **149**, 10–26.
- CHAMBERS, T. J., WEIR, R. C., GRAKOU, A., MCCOURT, D. W., BAZAN, J. F., FLETTERICK, R. J., and RICE, C. M. (1990). Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. *Proc. Natl. Acad. Sci. USA* **87**, 8898–8902.
- CLARKE, D. H., and CASALS, J. (1958). Techniques for hemagglutination and hemagglutination-inhibition with arthropod-borne viruses. *Am. J. Trop. Med. Hyg.* **7**, 561–573.
- COIA, G., PARKER, M. D., SPEIGHT, G., BYRNE, M. E., and WESTAWAY, E. G. (1988). Nucleotide and complete amino acid sequences of Kunjin virus: Definitive gene order and characteristics of the virus-specified proteins. *J. Gen. Virol.* **69**, 1–22.
- DALGARNO, L., TRENT, D. W., STRAUSS, J. H., and RICE, C. M. (1986). Partial nucleotide sequence of the Murray Valley encephalitis virus genome. Comparison of the encoded polypeptides with yellow fever virus structural and non-structural proteins. *J. Mol. Biol.* **187**, 309–323.
- DAYHOFF, M. O., SCHWARTZ, R. N., and ORCUTT, B. C. (1978). *In*



- "Atlas of Protein Sequence and Structure" (M. O. Dayhoff, Ed.), Vol. 5, Suppl. 3, p. 345. National Biomedical Research Foundation, Washington.
- DEUBEL, V., KINNEY, R. M., and 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., and 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.
- DEVEREUX, J., HAEBERLI, P., and SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387–395.
- DIAMOND, M. E., DOWHANICK, J. J., NEMEROFF, M. E., PIETRAS, D. F., TU, C., and BRUENN, J. A. (1989). Overlapping genes in a yeast double-stranded RNA virus. *J. Virol.* **63**, 3983–3990.
- DOJLA, V. V., BOYKO, V. P., AGRANOVSKY, A. A., and KOONIN, E. V. (1991). Phylogeny of capsid proteins of rod-shaped and filamentous RNA plant viruses: Two families with distinct patterns of sequence and probably structure conservation. *Virology* **184**, 79–86.
- DOMIER, L. L., FRANKLIN, K. M., SHAHABUDDIN, M., HELLMANN, G. M., OVERMEYER, J. H., HIREMATH, S. T., SIAW, M. F. E., LOMONOSOFF, G. P., SHAW, J. G., and RHOADS, R. E. (1986). The nucleotide sequence of tobacco vein mottling virus RNA. *Nucleic Acids Res.* **14**, 5417–5430.
- ECKELS, K. H., YONG-XIN, Y., DUBOIS, D. R., MARCHETTE, N. J., TRENT, D. W., and JOHNSON, A. J. (1988). Japanese encephalitis virus live-attenuated vaccine, Chinese strain SA<sub>14</sub>-14-2; Adaptation to primary canine kidney cell cultures and preparation of a vaccine for human use. *Vaccine* **6**, 513–518.
- FENG, D.-F., and DOOLITTLE, R. F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* **25**, 351–360.
- FENG, D.-F., and DOOLITTLE, R. F. (1990). Progressive alignment and phylogenetic tree construction of protein sequences. In "Methods in Enzymology" (R. F. Doolittle, Ed.), Vol. 183, pp. 375–387. Academic Press, San Diego.
- FENG, D.-F., JOHNSON, M. S., and DOOLITTLE, R. F. (1985). Aligning amino acid sequences: Comparison of commonly used methods. *J. Mol. Evol.* **21**, 112–125.
- GALLIONE, C. J., and ROSE, J. K. (1985). A single amino acid substitution in a hydrophobic domain causes temperature-sensitive cell-surface transport of a mutant viral glycoprotein. *J. Virol.* **54**, 374–382.
- GORBALENYA, A. E., and KOONIN, E. V. (1988). Birnavirus RNA polymerase is related to polymerases of positive strand RNA viruses. *Nucleic Acids Res.* **16**, 7735.
- GORBALENYA, A. E., KOONIN, E. V., DONCHENKO, A. P., and BLINOV, V. M. (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* **17**, 4713–4727.
- GOULD, E. A., BUCKLEY, A., BARRETT, A. D. T., and CAMMACK, N. (1986). Neutralizing (54K) and non-neutralizing (54K and 48K) monoclonal antibodies against structural and non-structural yellow fever virus proteins confer immunity in mice. *J. Gen. Virol.* **67**, 591–595.
- GUILLEY, H., CARRINGTON, J. C., BALAZS, E., JONARD, G., RICHARDS, K., and MORRIS, T. J. (1985). Nucleotide sequence and genome organization of carnation mottle virus RNA. *Nucleic Acids Res.* **13**, 6663–6677.
- HABILI, N., and SYMONS, R. H. (1989). Evolutionary relationship between luteoviruses and other RNA plant viruses based on sequence motifs in their putative RNA polymerases and nucleic acid helicases. *Nucleic Acids Res.* **17**, 9543–9555.
- HAHN, C. S., DALRYMPLE, J. M., STRAUSS, J. H., and RICE, C. M. (1987). Comparison of the virulent Asibi strain of yellow fever virus with the 17D vaccine strain derived from it. *Proc. Natl. Acad. Sci. USA* **84**, 2019–2023.
- HAHN, Y. S., GALLER, R., HUNKAPILLER, T., DALRYMPLE, J. M., STRAUSS, J. H., and STRAUSS, E. G. (1988). Nucleotide sequence of dengue 2 RNA and comparison of the encoded proteins with those of other flaviviruses. *Virology* **162**, 167–180.
- HALSTEAD, S. B. (1981). The pathogenesis of dengue: Molecular epidemiology in infectious disease. *Am. J. Epidemiol.* **114**, 632–648.
- HALSTEAD, S. B., and 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.
- HENCHAL, E. A., HENCHAL, L. S., and SCHLESINGER, J. J. (1988). Synergistic interactions of anti-NS1 monoclonal antibodies protect passively immunized mice from lethal challenge with dengue 2 virus. *J. Gen. Virol.* **69**, 2101–2107.
- HIGGINS, C. F., HILES, J. D., SALMOND, G. P. C., GILL, D. R., DOWNIE, J. A., EVANS, I. J., HOLLAND, I. B., GRAY, L., BUCKEL, S. D., BELL, A. W., and HERMODSON, M. A. (1986). A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. *Nature* **323**, 448–450.
- ICHO, T., and WICKNER, R. B. (1989). The double-stranded RNA genome of yeast virus L-A encodes its own putative RNA polymerase by fusing two open reading frames. *J. Biol. Chem.* **264**, 6716–6723.
- IGARASHI, A. (1978). Isolation of a Singh's *Aedes albopictus* cell clone sensitive to dengue and Chikungunya viruses. *J. Gen. Virol.* **40**, 531–544.
- INNIS, B. L., THIRAWUTH, V., and HEMACHUDHA, C. (1989). Identification of continuous epitopes of the envelope glycoprotein of dengue type 2 virus. *Am. J. Trop. Med. Hyg.* **40**, 676–687.
- IRIE, K., MOHAN, P. M., SASAGURI, Y., PUTNAK, R., and PADMANABHAN, R. (1989). Sequence analysis of cloned dengue virus type 2 genome (New Guinea-C strain). *Gene* **75**, 197–211.
- KAMER, G., and ARGOS, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res.* **12**, 7269–7282.
- KAUFMAN, B. M., SUMMERS, P. L., DUBOIS, D. R., and ECKELS, K. H. (1987). Monoclonal antibodies against the dengue 2 virus E-glycoprotein protect mice against lethal dengue infection. *Am. J. Trop. Med. Hyg.* **36**, 427–434.
- KAUFMAN, B. M., SUMMERS, P. L., DUBOIS, D. R., COHEN, W. H., GENTRY, M. K., TIMCHAK, R. L., BURKE, D. S., and ECKELS, K. H. (1989). Monoclonal antibodies for dengue virus prM glycoprotein protect mice against lethal dengue infection. *Am. J. Trop. Med. Hyg.* **41**, 576–580.
- KOONIN, E. V., GORBALENYA, A. E., and CHUMAKOV, K. M. (1989). Tentative identification of RNA-dependent RNA polymerases of dsRNA viruses and their relationship to positive strand RNA viral polymerases. *FEBS Lett.* **252**, 42–46.
- LAI, C.-J., ZHAO, B., HORI, H., and BRAY, M. (1991). Infectious RNA transcribed from stably cloned full-length cDNA of dengue type 4 virus. *Proc. Natl. Acad. Sci. USA* **88**, 5139–5143.
- LAIN, S., RIECHMANN, J. L., MARTIN, M. T., and GARCIA, J. A. (1989). Homologous potyvirus and flavivirus proteins belonging to a superfamily of helicase-like proteins. *Gene* **82**, 357–362.
- MACKOW, E., MAKINO, Y., ZHAO, B., ZHANG, Y.-M., MARKOFF, L., BUCKLER-WHITE, A., GUILER, M., CHANOCK, R., and LAI, C.-J. (1987). The nucleotide sequence of dengue type 4 virus: Analysis of genes coding for nonstructural proteins. *Virology* **159**, 217–228.

- MANDL, C. W., HEINZ, F. X., and KUNZ, C. (1988). Sequence of the structural proteins of tick-borne encephalitis virus (Western subtype) and comparative analysis with other flaviviruses. *Virology*, **166**, 197–205.
- MANDL, C. W., GUIRAKHO, F., HOLZMANN, H., HEINZ, F. X., and KUNZ, C. (1989a). Antigenic structure of the flavivirus envelope protein E at the molecular level, using tick-borne encephalitis virus as a model. *J. Virol.* **63**, 564–571.
- MANDL, C. W., HEINZ, F. X., STÖCKL, E., and KUNZ, C. (1989b). Genome sequence of tick-borne encephalitis virus (Western subtype) and comparative analysis of nonstructural proteins with other flaviviruses. *Virology* **173**, 291–301.
- MANIATIS, T., FRITSCH, E. F., and SAMBROOK, J. (1982). "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- MASON, P. W., McADA, P. C., MASON, T. L., and FOURNIER, M. J. (1987). Sequence of the dengue-1 virus genome in the region encoding the three structural proteins and the major nonstructural protein NS1. *Virology* **161**, 262–267.
- McADA, P. C., MASON, P. W., SCHMALJOHN, C. S., DALRYMPLE, J. M., MASON, T. L., and FOURNIER, M. J. (1987). Partial nucleotide sequence of the Japanese encephalitis virus genome. *Virology* **158**, 348–360.
- McCLURE, M. A., JOHNSON, M. S., FENG, D.-F., and DOOLITTLE, R. F. (1988). Sequence comparisons of retroviral proteins: Relative rates of change and general phylogeny. *Proc. Natl. Acad. Sci. USA* **85**, 2469–2473.
- MILLER, R. H., and PURCELL, R. H. (1990). Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. *Proc. Natl. Acad. Sci. USA* **87**, 2057–2061.
- NEEDLEMAN, S. B., and WUNSCH, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–453.
- NITAYAPHAN, S., GRANT, J. A., CHANG, G. J., and TRENT, D. W. (1990). Nucleotide sequence of the virulent SA-14 strain of Japanese encephalitis virus and its attenuated vaccine derivative, SA-14-14-2. *Virology* **177**, 541–552.
- NOWAK, T., and WENGLER, G. (1987). Analysis of disulfides present in the membrane proteins of the West Nile flavivirus. *Virology* **156**, 127–137.
- OKAYAMA, H., and BERG, P. (1982). High-efficiency cloning of full-length cDNA. *Mol. Cell. Biol.* **2**, 161–170.
- OSATOMI, K., and SUMIYOSHI, H. (1990). Complete nucleotide sequence of dengue type 3 virus genome RNA. *Virology* **176**, 643–647.
- PIETRAS, D. F., DIAMOND, M. E., and BRUENN, J. A. (1988). Identification of a putative RNA dependent RNA polymerase encoded by a yeast double stranded RNA virus. *Nucleic Acids Res.* **16**, 6225.
- POCH, O., SAUVAGET, I., DELARUE, M., and TORDO, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8**, 3867–3874.
- RICE, C. M., GRAKOU, A., GALLER, R., and CHAMBERS, T. J. (1989). Transcription of infectious yellow fever virus RNA from full-length cDNA templates produced by *in vitro* ligation. *New Biologist* **1**, 285–296.
- RICE, C. M., LENCHES, E. M., EDDY, S. R., SHIN, S. J., SHEETS, R. L., and STRAUSS, J. H. (1985). Nucleotide sequence of yellow fever virus: Implications for flavivirus gene expression and evolution. *Science* **229**, 726–733.
- RICE, C. M., STRAUSS, E. G., and STRAUSS, J. H. (1986). Structure of the flavivirus genome. In "Togaviruses and Flaviviruses" (S. Schlesinger and M. Schlesinger, Eds.), pp. 279–327. Plenum, New York.
- ROBAGLIA, C., DURAND-TARDIF, M., TRONCHET, M., BOUDAZIN, G., ASTIER-MANIFACIER, S., and CASSE-DELBART, F. (1989). Nucleotide sequence of potato virus Y (N strain) genomic RNA. *J. Gen. Virol.* **70**, 935–947.
- ROSEN, L. (1977). *The Emperor's New Clothes* revisited, or reflections on the pathogenesis of dengue hemorrhagic fever. *Am. J. Trop. Hyg. Med.* **26**, 337–343.
- ROSSMANN, M. G., and RUECKERT, R. R. (1987). What does the molecular structure of viruses tell us about viral functions? *Microbiol. Sci.* **4**, 206–214.
- SAITOU, N., and NEI, M. (1987). The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
- SANGER, F., NICKLEN, S., and COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- SCHLESINGER, J. J., BRANDISS, M. W., and WALSH, E. E. (1985). Protection against 17D yellow fever encephalitis in mice by passive transfer of monoclonal antibodies to the nonstructural glycoprotein GP48 and by active immunization with GP48. *J. Immunol.* **135**, 2805–2809.
- SCHLESINGER, J. J., BRANDISS, M. W., and WALSH, E. E. (1987). Protection of mice against dengue 2 virus encephalitis by immunization with the dengue 2 virus non-structural glycoprotein NS1. *J. Gen. Virol.* **68**, 853–857.
- SCHLESINGER, J. J., BRANDISS, M. W., CROPP, C. P., and MONATH, T. P. (1986). Protection against yellow fever in monkeys by immunization with yellow fever virus nonstructural protein NS1. *J. Virol.* **60**, 1153–1155.
- SCHLESINGER, R. W. (1977). Dengue viruses. In "Virology Monographs," (S. Gard and C. Hallaver, Eds.), Vol. 16. Springer-Verlag, Wien, New York.
- SPEIGHT, G., COIA, G., PARKER, M. D., and WESTAWAY, E. G. (1988). Gene mapping and positive identification of the nonstructural proteins NS2A, NS2B, NS3, NS4B and NS5 of the flavivirus Kunjin and their cleavage sites. *J. Gen. Virol.* **69**, 23–34.
- STUDIER, J. A., and KEPPLER, K. J. (1988). A note on the neighbour-joining tree algorithm of Saitou and Nei. *Mol. Biol. Evol.* **5**, 729–731.
- SUMIYOSHI, H., MORI, C., FUKU, I., MORITA, K., KUHARA, S., KONDOU, J., KIKUCHI, Y., NAGAMATU, H., and IGARASHI, A. (1987). Complete nucleotide sequence of the Japanese encephalitis virus genome RNA. *Virology* **161**, 497–510.
- TEYCHENEY, P. Y., TAVER, G., DELBOS, R., RAVELONANDRO, M., and DUNEZ, J. (1989). The complete nucleotide sequence of plum pox virus RNA (strain D). *Nucleic Acids Res.* **17**, 10,115–10,116.
- THEILER, M., and SMITH, H. H. (1937). Use of yellow fever virus modified by *in vitro* cultivation for human immunization. *J. Exp. Med.* **65**, 787–800.
- TRENT, D. W., KINNEY, R. M., JOHNSON, B. J. B., VORNDAM, A. V., GRANT, J. A., DEUBEL, V., RICE, C. M., and HAHN, C. (1987). Partial nucleotide sequence of St. Louis encephalitis virus RNA: Structural proteins, NS1 ns2a, and ns2b. *Virology* **156**, 293–304.
- VANDEPOL, S. B., and HOLLAND, J. J. (1986). Evolution of vesicular stomatitis virus in athymic nude mice: Mutations associated with natural killer cell selection. *J. Gen. Virol.* **67**, 441–451.
- WALKER, J. E., SARASTE, M., RUNSWICK, M. J., and GAY, N. J. (1982). Distantly related sequences in the  $\alpha$  and  $\beta$  subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945–951.

- WENGLER, G., and WENGLER, G. (1981). Terminal sequences of the genome and replicative form RNA of the flavivirus West Nile virus: Absence of the poly (A) and possible role in RNA replication. *Virology* **113**, 544–555.
- WENGLER, G., CASTLE, E., LEIDNER, U., NOWAK, T., and WENGLER, G. (1985). Sequence analysis of the membrane protein V3 of the flavivirus West Nile virus and of its gene. *Virology* **147**, 264–274.
- WESTAWAY, E. G., BRINTON, M. A., GAIDAMOVICH, S. YA., HORZINEK, M. C., IGARASHI, A., KÄÄRIÄINEN, L., LVOV, D. K., PORTERFIELD, J. S., RUSSELL, P. K., and TRENT, D. W. (1985). Flaviviridae. *Intervirology* **24**, 183–192.
- YAMSHCHIKOV, V. F., and PLETNEV, A. G. (1988). Nucleotide sequence of the genome region encoding the structural proteins and the NS1 protein of the tick borne encephalitis virus. *Nucleic Acids Res.* **16**, 7750.
- YOKSAN, S., BHAMARAPRAVATI, N., and HALSTEAD, S. B. (1986). Dengue virus vaccine development: Study on biological markers of uncloned dengue 1-4 viruses serially passaged in primary kidney cells. In "Arbovirus Research in Australia" Proc Fourth Symp. (T. D. St. George, B. H. Kay, and J. Blok, Eds.), pp. 35–38. CSIRO/QIMR, Brisbane.
- ZHANG, Y.-M., HAYES, E. P., MCCARTY, T. C., DUBOIS, D. R., SUMMERS, P. L., ECKELS, K. H., CHANOCK, R. M., and LAI, C.-J. (1988). Immunization of mice with dengue structural proteins and nonstructural protein NS1 expressed by baculovirus recombinant induces resistance to dengue virus encephalitis. *J. Virol.* **62**, 3027–3031.
- ZHAO, B., MACKOW, E., BUCKLER-WHITE, A., MARKOFF, L., CHANOCK, R. M., LAI, C.-J., and MAKINO, Y. (1986). Cloning full-length dengue 4 viral DNA sequences: Analysis of genes coding for structural proteins. *Virology* **155**, 77–88.