Journal of General Virology (1997), **78**, 2293–2297. Printed in Great Britain

Extensive nucleotide changes and deletions within the envelope glycoprotein gene of Euro-African West Nile viruses

François-Xavier Berthet,[1]† Hervé G. Zeller,[2] Marie-Thérèse Drouet,[3] Jean Rauzier,[3]† Jean-Pierre Digoutte[1] and Vincent Deubel[3]

[1] Institut Pasteur, BP 220, Dakar, Senegal

[2] Institut Pasteur de Madagascar, BP 1724, Antananarivo, Madagascar

[3] Unité des Arbovirus et Virus des Fièvres Hémorragiques, Institut Pasteur-Paris, 25 rue du Dr Roux, 75724 Paris cedex 15, France

We compared the sequence of an envelope protein gene fragment from 21 temporally distinct West Nile (WN) virus strains, isolated in nine African countries and in France. Alignment of nucleotide sequences defined two groups of viruses which diverged by up to 29%. The first group of subtypes is composed of nine WN strains from France and Africa. The Austral-Asian Kunjin virus was classified as a WN subtype in this first group. The second group includes 12 WN strains from Africa and Madagascar. Four strains harboured a 12 nucleotide in-frame deletion. The loss of the corresponding four amino acids resulted in the loss of the potential glycosylation site present in several WN strains. The distribution of virus subtypes into two lineages did not correlate with host preference or geographical origin. The isolation of closely related subtypes in distant countries is consistent with WN viruses being disseminated by migrating birds.

West Nile (WN) virus is a flavivirus (family *Flaviviridae*) which shares antigenic properties with other members of the Japanese encephalitis (JE) virus serogroup including JE, St Louis encephalitis (SLE), Murray Valley encephalitis (MVE) and Kunjin viruses (Monath & Heinz, 1996). WN virus is a mosquito-borne pathogen responsible for WN fever in humans (Hayes, 1988). Clinical symptoms associated with WN virus infection are most frequently those of a mild febrile illness, but fatal cases of acute meningoencephalitis and fulminant hepatitis have been reported (Monath & Heinz, 1996). WN virus is endemic in tropical areas, particularly India and Africa, where local proliferation of infected mosquitoes can lead to epizootics (Madagascar in 1982 and Senegal in 1988 and 1990) and sudden epidemics (Israel in 1950 and South Africa in 1974). Recent epidemics with high rates of patient mortality have been reported in Algeria (1995) and in Romania (Anon., 1996; Le Guenno *et al.*, 1996) presenting WN virus as a potential emerging human-pathogenic virus. WN virus can infect a wide range of vertebrate species in nature (Hayes, 1988) and wild birds are believed to play an important role in the WN transmission cycle by disseminating the virus during migration (Hayes, 1988; Monath & Heinz, 1996). Because WN virus has a wide geographical distribution, regional variation was investigated from an early stage. Thus, Hammam *et al.* (1965) found that Indian and African WN virus isolates have different haemagglutination inhibition kinetics. Studies of cDNA/RNA heteroduplex restriction profiles and/or reactivity toward monoclonal antibodies (Besselaar & Blackburn, 1988; Mathiot *et al.*, 1990) have identified several WN virus variants. More recently, Porter *et al.* (1993) sequenced NS3 protein gene fragments of seven African and one Indian strains and defined three categories on the basis of nucleotide sequence similarity.

In this study, we report the classification of 21 WN strains (Table 1) isolated in Africa and France by direct nucleotide sequencing of a gene fragment obtained from genomic RNA by RT–PCR. Cytoplasmic RNA was extracted from virus-infected *Aedes pseudoscutellaris* AP61 cells using a previously described method (Deubel *et al.*, 1993). Nucleic acid sequence analysis was performed on an envelope (E) gene fragment obtained by RT–PCR from positions 1318 to 1645 in the WN virus genome (Wengler *et al.*, 1985). The choice was dictated by the variability of this region as demonstrated by comparative amino acid sequence analysis in the flavivirus group. In particular, four amino acids at positions 154 to 157 in the E protein of Kunjin virus (the closest virus to WN virus in the flavivirus evolutionary tree) are absent from the published sequence of a Nigerian WN virus (Wengler *et al.*, 1985).

Author for correspondence: Vincent Deubel.
Fax +33 1 45 68 87 80. e-mail vdeubel@pasteur.fr

† **Present address:** Unité de Génétique Mycobactérienne, Institut Pasteur-Paris, 25 rue du Dr Roux, 75724 Paris cedex 15, France.

The nucleotide sequence data reported in this paper have been deposited in the GenBank database and assigned accession numbers AF001556–AF001574.

Table 1. Characteristics of the 21 West Nile strains analysed in this study

WN strains were obtained from the WHO collaborating centres for reference and research on arboviruses (Pasteur Institutes in Paris and Dakar). Virus strains were recovered from and/or passaged on either suckling mouse brain (before 1990) or cell culture (after 1990).

Strain	Geographical origin	Year of isolation	Primary source of isolation†
Na1047	Kenya	Unknown	(M) *Aedes albothorax*
EntM63134	Uganda	Unknown	Unknown
MP22	Uganda	Unknown	Unknown
HEg101	Egypt	1951	Human
PaH651	France	1965	Human
ArAlg/Djanet	Algeria	1968	(M) *Culex* sp.
ArB310	Central African Republic	1967	(M) *Culex* sp.
AnB3507	Central African Republic	1972	(B) *Antichromus minutus*
ArB3573	Central African Republic	1972	(M) *Culex tigripes*
HB83P55	Central African Republic	1983	Human
HB6343	Central African Republic	1989	Human
AnMg798	Madagascar	1978	(B) *Coracopsis vasa*
ArMg956	Madagascar	1986	(M) *Culex quinquefasciatus*
ArMg978	Madagascar	1988	(M) *Culex univittatus*
AnD27875	Senegal	1979	(P) *Galago senegalensis*
ArD76104*	Senegal	1990	(M) *Mimomyia lacustris*
ArD76986	Senegal	1990	(M) *Culex poicilipes*
ArD78016*	Senegal	1990	(M) *Aedes vexans*
ArD93548	Senegal	1993	(M) *Culex neavei*
ArA3212	Ivory Coast	1981	(M) *Culex guiarti*

* Sequences of the two strains were confirmed using cDNA products obtained by direct amplification of virus RNA in mosquito lysate pools.
† Mosquito (M), bird (B) or primate (P) species.

Moreover, preliminary studies of WN E protein by Western blotting suggested a variability in E protein glycosylation status similar to that observed among Kunjin isolates (V. Deubel, R. Hall & J. Mackenzie, unpublished results; Adams *et al.*, 1995).

cDNA was synthesized from WN virus RNAs using primer WN240 (5′ GAGGTTCTTCAAACTCCAT 3′) and amplified using primers WN240 and WN132 (5′ GAAAACATCAA-GTATGAGG 3′) as described previously (Deubel *et al.*, 1993). Primers WN132 and WN240 correspond to highly conserved sequences in viruses of the JE virus serogroup. DNA amplicons were purified by ion exchange chromatography and precipitated with 2 vols of isopropanol. Each amplicon (1 pmol) was mixed with 1·0 pmol of primer (WN132 or WN240) and sequenced using the *Taq* Dye Deoxy Terminator cycle sequencing kit (Applied Biosystems). Nucleic acid sequences were obtained on an automated Applied Biosystems 373A sequencer. WN virus sequences were aligned with each other using the multiple sequence alignment software CLUSTAL V (Higgins & Sharp, 1988).

The limited sequencing of nucleotides 436 to 690 in the E gene indicated a generally uniform rate of random nucleotide mutations (data not shown). However, four WN strains, including the Nigerian isolate described by Wengler *et al.* (1985), a strain from Uganda of unknown origin and two strains from Senegal isolated in 1990, showed a deletion of 12 nucleotides (nt 462 to 473). The sequenced regions of three of the strains were 100% identical. Each strain was separately cultured in cells and sequences of the two Senegalese strains were identified in RNA extracted from the original mosquito pool, which was preserved at −70 °C, thus excluding the possibility of laboratory contamination (see Table 1). The maximum nucleotide divergence in the sequenced region was 29%. Nucleic acid changes occurred at 99 positions (32% of the gene fragment); 69% of them were in the third codon and 85% were silent.

To visualize the range of relationships among WN viruses, a phylogram was constructed from the nucleotide sequences obtained in this study using the neighbour-joining method (Saitou & Nei, 1987). Because of its similarity to WN virus species, Kunjin virus was also included. The strains formed two distinct lineages (I and II) of closely related subtypes circulating in large and overlapping geographic areas (Fig. 1). The branches leading to these two lineages showed bootstrap values of 1000 and 954 when 1000 trees were sampled, indicating the robustness of the groupings. The lengths of the branches on the phylogenetic tree were proportional to evolutionary distance. Nucleotide sequences of virus subtypes

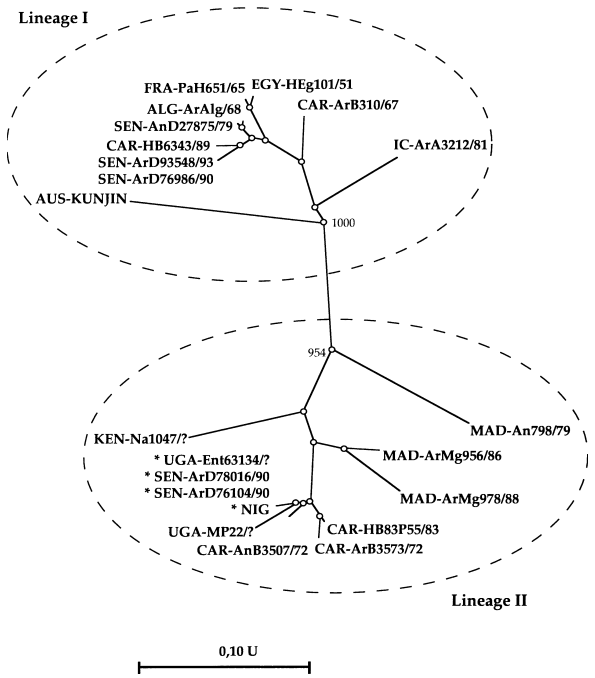

Fig. 1. Unrooted tree presenting the extent of nucleotide sequence identity from the E gene fragment of 21 WN viruses and Kunjin virus (Coia *et al*., 1988). Nucleotide sequences have been deposited in the GenBank database: SEN-ArD78016, AF001556; CAR-HB83P55, AF001557; CAR-HB6343, AF001558; MAD-AnMg798, AF001559; FRA-PaH651, AF001560; IC-ArA3212, AF001561; UGA-MP22, AF001562; RCA-AnB3507, AF001563; MAD-ArMg956, AF001564; RCA-ArB3573, AF001565; CAR-ArB310, AF001566; ALG-ArDjanet, AF001567; EGY-HEg101, AF001568; SEN-AnD27875, AF001569; SEN-ArD93548, AF001570; KEN-Na1047, AF001571; AUS-Kunjin, AF001572; UGA-Ent63134, AF001573; MAD-ArMg978, AF001574. The geographical origin, strain number and year of isolation are given for each isolate in Table 1. Asterisks indicate the four strains with 12 nucleotide deletions in the E gene fragment. The phylogenetic tree was constructed using the neighbour-joining method (Saitou & Nei, 1987) using the CLUSTAL V program (Higgins & Sharp, 1988). Bootstrap confidence limits between the two defined lineages I and II were calculated from 1000 replicate trees. The horizontal bar indicates the distance corresponding to 10% of nucleotide changes.

in one lineage differed by a maximum of 29% from those of subtypes in the second lineage. Within lineage I the maximum identity of WN strains was 87% and within lineage II it was 80·5%. Kunjin virus shared more than 80% nucleotide identity with WN viruses of lineage I and was consequently classified as a subtype in this lineage. Lineage I is composed of nine WN virus strains from France and North (Algeria, Egypt), West (Senegal, Ivory Coast) and Central (Central African Republic, CAR) Africa. Lineage II includes 12 WN virus strains from West (Senegal, Nigeria), Central (CAR) and East (Uganda, Kenya) Africa and Madagascar. Strains ArD76986 and ArD93548 isolated in Senegal in 1990 and 1993, respectively, have the same E gene fragment sequence. In contrast, strains ArD76104 and ArD78016 isolated in Senegal in 1990 differ from ArD76986, showing that very different subtypes circulate simultaneously in the same country. Three strains from Madagascar were isolated in different regions of the island and from different vectors and hosts. Strain AnMg798, isolated from a parrot, was classified in lineage II with strains ArMg978 and ArMg956 isolated from *Culex* species 8 to 10 years later, but showed more than 19% nucleotide divergence. In lineage II, strains with a 12 nucleotide deletion in the E gene fragment (indicated by an asterisk in Fig. 1) showed less than 5% nucleotide divergence from subtypes from CAR.

Amino acid differences were identified at 15 positions. Alignment of amino acids 146 to 230 of WN virus strains with that of the 1951 Egyptian HEg101 prototype WN strain revealed a maximum divergence of 13% (Fig. 2). Amino acids 154 to 157 were absent from strains ArD76104, ArD78016, EntM63134 and Nigeria. The sequence at these positions in the other strains was NYST, NYPT or SYST. The triad NYS forms an *N*-glycosylation site at position 154 in the E protein in JE, MVE and SLE viruses, and in some Kunjin viruses (Adams *et al*., 1995). This confirms the previous observation that the E protein of French 1965 WN isolates (V. Deubel, R. Hall & J. Mackenzie, unpublished results) and that of the Sarafend WN strain (Ng *et al*., 1994) are glycosylated. However, we cannot exclude the possibility that passaging may have influenced the glycosylation status of the E protein of WN strains studied as described for Kunjin virus (Adams *et al*., 1995). Direct sequencing of cDNA amplified from virus RNA in mosquito pools or in patient sera would confirm the polymorphism of the glycosylation site in natural infecting viruses. Clear signature motifs at amino acid positions A → S 172, N → S 199, T → S 205, T → A 208 and T → S 210 confirmed WN-Kunjin genotype distribution into two lineages. The corresponding published sequence from Kunjin showed only 4 to 11 amino acid changes (5 to 13% divergence) with WN strains, whereas other viruses from the JE serogroup exhibited 24 to 38 amino acid changes (28 to 45% divergence). This result confirmed the close relatedness of Kunjin virus and WN viruses (Coia *et al*., 1988).

Little information about the genetic diversity of WN viruses was available before this study. Comparative serological analysis of WN strains indicated a variability between isolates from different hosts and vectors and geographic areas but also showed that different strains circulated in the same area at the same time (Gaidamovich & Sokhey, 1973; Hammam *et al*., 1965; Hammam & Price, 1966; Mathiot *et al*., 1990; Odelola & Fabiyi, 1976; Price & O'Learly, 1967; Umrigar & Pavri, 1977). We previously demonstrated that the Egyptian prototype HEg101 and the Nigerian strains diverge by 22% in the nucleotide sequences corresponding to the C terminus of NS5 and to part of the 3′ non coding region (Pierre *et al*., 1994). Porter *et al*. (1993) compared nucleotide sequences at 182 positions in the NS3 gene of eight WN strains from seven countries and reported closer relationships between the Nigerian and Ugandan strains (99·5% similarity) and to a lesser extent with Malagasy strain (86% identity), than with strains from CAR, Ethiopia, Egypt and India which shared more than

```
                  146      *
EGY-HEg101/51     PTTVESHGNYSTQIGATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFWVHREWFMDLNLPWSSAGS
FRA-PaH651/65     ................................................................................
ALG-ArAlg/68      .........P.....................................................L.................
SEN-AnD27875/79   .........P.....................................................L.................
SEN-ArD76986/90   ................................................................L.................
SEN-ArD93548/93   ................................................................L.................
CAR-HB6343/89     .........P.............S.....................M.............S.L.................
CAR-ArB310/67     .........P.....................................................L.................
IC-ArA3212/81     .........P.....................................S...........S.L.................

AUS-KUNJIN        .........F..T..A.....................................S...........L.............E.

SEN-ArD76104/90   ........----K.............S.........................S.....S..A.S.L.................
SEN-ArD78016/90   ........----K.............S.........................S.....S..A.S.L.................
UGA-EntM63134/?   ........----K.............S.........................S.....S..A.S.L.................
NIG               ........----K.............S.........................S.....S..E.S.L.................
CAR-AnB3507/72    ........................S.........................S.....S..A.S.L.................
CAR-ArB3573/72    ........................S.........................S.....S..A.S.L.................
CAR-HB83P55/83    ........................S.........................S.....S..A.S.L.................
UGA-MP22/?        ........S...............S.........................S.....S..A.S.L.................
MAD-ArMg956/86    ............V...........S.........D...............S.....S..A.S.L.................
MAD-ArMg978/88    ........................S.........................S.....S..A.S.L.................
KEN-Na1047/?      ............V...........S.........................S.....S..A.S.L.............G
MAD-AnMg798/78    ............V...........S.....M...D.....I.........S.....S..A.S.L.............G

JE                T..S.N.....A.V..S..AK.TV..N...I.....D.....L.......LN.E.F......S.S.L......H..A...TPPS.
MVE               S.DST..........N..V..T.S.N..AI.A.M.D......E......LN.E......I...H.L......N..L...T.PA.
SLE               S.DST......E...KN..A..T.S.Q...F.ANM....T..I...A....N.ED...F..KE.SWL.N.D..H.....T.PAT
```

Fig. 2. Amino acid sequences deduced from the 255 nucleotides used to determine genetic relatedness among 21 WN and one Kunjin isolates. Sequences of the Nigerian WN (Wengler *et al*., 1985), Kunjin (Coia *et al*., 1988), JE, MVE and SLE (collected in Trent *et al*., 1987) viruses are aligned. The first amino acid characterized in the WN E protein sequence (Wengler *et al*., 1985) is indicated. Dots indicate amino acid identities with the Egyptian HEg101 WN and dashes show missing amino acids. The potential *N*-glycosylation site is indicated by an asterisk.

92% identity but diverged by about 24% from the first grouping. Our database on the nucleotide sequence of an E gene fragment confirm this classification and emphasize the substantial genetic variability among African strains, which are unrelated to geographical region, time and host/vector.

Comparative genetic analysis of other flaviviruses including dengue, yellow fever (YF), tick-borne encephalitis (TBE) and JE viruses showed a similar degree of variability within each species, although the classification of TBE and YF viruses correlates with geographic origin and corresponds to topotypes (Zanotto *et al*., 1996). However, dengue viruses are carried from one continent to the other by viraemic patients travelling by air. Like JE virus, WN virus uses a variety of hosts and vectors for its maintenance and propagation in nature. Wild birds were quickly recognized as important vertebrate hosts for WN virus and their migrations are assumed to be instrumental in virus dissemination (Taylor *et al*., 1956). Thus, similar subtypes found in France and Senegal and in Senegal and Central Africa suggest continent-to-continent and land-to-land movements. The involvement of birds in the repeated import of WN strains may explain the multiplicity of subtypes in Senegal and in CAR. The data presented in Fig. 1 strongly support the idea that these viruses are free to move between regions and organisms, with no obvious selection.

Sequencing a short fragment of the E gene of WN viruses responsible for recent epidemics, and strains from South Africa and from other parts of the world (former Soviet Union, Israel, Pakistan and India) would give an overall picture of strain variability and virus circulation and may bring new insights into possible relationships with phenotypic variability.

We are grateful to Aïcha Diouf, Annick Leclerc and Mireille Mondo for help in cell culture and virus propagation procedures. We thank Michèle Bouloy and Bahloul Chokri for help in phylogenetic tree analysis and Dr Brigitte Gicquel for providing automated DNA sequencing facilities. François-Xavier Berthet was supported by a doctoral grant from the Agence Nationale de Recherche sur le SIDA (ANRS).

References

Adams, S. C., Broom, A. K., Sammels, L. M., Harnett, A. C., Howard, M. J., Coelen, R. J., Mackenzie, J. S. & Hall, R. A. (1995). Glycosylation and antigenic variation among Kunjin virus isolates. *Virology* **206**, 49–56.

Anon. (1996). Viral meningitidis. *Weekly Epidemiological Record* **71**, 296.

Besselaar, T. G. & Blackburn, N. K. (1988). Antigenic analysis of West Nile virus strains using monoclonal antibodies. *Archives of Virology* **99**, 75–88.

Coia, G., Parker, M. D., Speight, G., Byrne, M. E. & Westaway, E. G. (1988). Nucleotide and complete amino acid sequences of Kunjin virus: definitive gene order and characteristics of the virus-specified proteins. *Journal of General Virology* **69**, 1–21.

Deubel, V., Nogueira, R. M., Drouet, M. T., Zeller, H., Reynes, J. M. & Ha, D. Q. (1993). Direct sequencing of genomic cDNA fragments amplified by the polymerase chain reaction for molecular epidemiology of dengue-2 viruses. *Archives of Virology* **129**, 197–210.

Gaidamovich, S. Y. & Sokhey, J. (1973). Studies on antigenic peculiarities of West Nile virus strains isolated in the USSR by three serological tests. *Acta Virologica* **17**, 343.

Hammam, H. M. & Price, W. H. (1966). Further observations on geographic variation in the antigenic character of West Nile and Japanese B encephalitis viruses. *American Journal of Epidemiology* **83**, 113.

Hammam, H. M., Clark, D. H. & Price, W. H. (1965). Antigenic variations of West Nile virus in relation to geography. *American Journal of Epidemiology* **82**, 40–55.

Hayes, C. G. (1988). West Nile Fever. In *The Arboviruses: Epidemiology and Ecology*, pp. 59–89. Edited by T. P. Monath. Boca Raton, Fla.: CRC Press.

Higgins, D. G. & Sharp, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**, 237–244.

Le Guenno, B., Bougermouth, A., Dif, A., Azzam, T. & Bouakaz, R. (1996). West Nile, a deadly virus? *Lancet* **348**, 1315.

Mathiot, C. C., Georges, A. J. & Deubel, V. (1990). Comparative analysis of West Nile virus strains isolated from human and animal hosts

using monoclonal antibodies and cDNA restriction digest profiles. *Research in Virology* **141**, 533–543.

Monath, T. P. & Heinz, F. X. (1996). Flaviviruses. In *Virology*, pp. 961–1034. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia: Lippincott–Raven.

Ng, M. L., Howe, J., Sreenivasan, V. & Mulders, J. J. L. (1994). Flavivirus West Nile (Sarafend) egress at the plasma membrane. *Archives of Virology* **137**, 303–313.

Odelola, H. A. & Fabiyi, A. (1976). Antigenic relationships among Nigerian strains of West Nile virus by complement fixation and agar gel precipitation techniques. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **70**, 138.

Pierre, V., Drouet, M. T. & Deubel, V. (1994). Identification of mosquito-borne flavivirus sequences using universal primers and reverse transcription/polymerase chain reaction. *Research in Virology* **145**, 93–104.

Porter, K. R., Summers, P. L., Dubois, D., Puri, B., Nelson, W., Henchal, E., Oprandy, J. J. & Hayes, C. G. (1993). Detection of West Nile virus by the polymerase chain reaction and analysis of nucleotide sequence variation. *American Journal of Tropical Medicine and Hygiene* **48**, 440–446.

Price, W. H. & O'Learly, W. (1967). Geographic variation in the antigenic character of West Nile virus. *American Journal of Epidemiology* **85**, 84–86.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular and Biological Evolution* **4**, 406–425.

Taylor, R. M., Work, T. H., Hurlbut, H. S. & Zizk, F. (1956). A study of the ecology of West Nile virus in Egypt. *American Journal of Tropical Medicine and Hygiene* **5**, 579–620.

Trent, D. W., Kinney, R. M., Johnson, B. J., Vorndam, A. V., Grant, J. A., Deubel, V., Rice, C. M. & Hahn, C. S. (1987). Partial nucleotide sequence of St. Louis encephalitis virus RNA: structural proteins, NS1, ns2a and ns2b. *Virology* **256**, 293–304.

Umrigar, M. D. & Pavri, K. K. M. (1977). Comparative serological studies on Indian strains of West Nile virus isolated from different sources. *Indian Journal of Medical Research* **65**, 603.

Wengler, G., Castle, E., Leidner, U., Nowak, T. & Wengler, G. (1985). Sequence analysis of the membrane protein V3 of the flavivirus West Nile virus and of its gene. *Virology* **147**, 264–274.

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

Received 6 March 1997; Accepted 14 May 1997