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 Digoutte1 and Vincent Deubel3

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 pseudocubellaris 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).
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).

<table>
<thead>
<tr>
<th>Strain</th>
<th>Geographical origin</th>
<th>Year of isolation</th>
<th>Primary source of isolation†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na1047</td>
<td>Kenya</td>
<td>Unknown</td>
<td>(M) Aedes albopictus</td>
</tr>
<tr>
<td>EntM63134</td>
<td>Uganda</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>MP22</td>
<td>Uganda</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>HEg101</td>
<td>Egypt</td>
<td>1951</td>
<td>Human</td>
</tr>
<tr>
<td>PaH651</td>
<td>France</td>
<td>1965</td>
<td>Human</td>
</tr>
<tr>
<td>ArAlg/Djanet</td>
<td>Algeria</td>
<td>1968</td>
<td>(M) Culex sp.</td>
</tr>
<tr>
<td>AnB3507</td>
<td>Central African Republic</td>
<td>1972</td>
<td>(B) Antichromus minutus</td>
</tr>
<tr>
<td>ArB3573</td>
<td>Central African Republic</td>
<td>1972</td>
<td>(M) Culex tigripes</td>
</tr>
<tr>
<td>HB83P55</td>
<td>Central African Republic</td>
<td>1983</td>
<td>Human</td>
</tr>
<tr>
<td>HB6343</td>
<td>Central African Republic</td>
<td>1989</td>
<td>Human</td>
</tr>
<tr>
<td>ArMg798</td>
<td>Madagascar</td>
<td>1978</td>
<td>(B) Coracopsis vasa</td>
</tr>
<tr>
<td>ArMg956</td>
<td>Madagascar</td>
<td>1986</td>
<td>(M) Culex quinquefasciatus</td>
</tr>
<tr>
<td>ArMg978</td>
<td>Madagascar</td>
<td>1988</td>
<td>(M) Culex univittatus</td>
</tr>
<tr>
<td>AnD768775</td>
<td>Senegal</td>
<td>1979</td>
<td>(P) Galago senegalensis</td>
</tr>
<tr>
<td>ArD76104*</td>
<td>Senegal</td>
<td>1990</td>
<td>(M) Mimomyia lacustris</td>
</tr>
<tr>
<td>ArD76986</td>
<td>Senegal</td>
<td>1990</td>
<td>(M) Culex poicilpes</td>
</tr>
<tr>
<td>ArD78016*</td>
<td>Senegal</td>
<td>1990</td>
<td>(M) Aedes vexans</td>
</tr>
<tr>
<td>ArD93548</td>
<td>Senegal</td>
<td>1993</td>
<td>(M) Culex neavei</td>
</tr>
<tr>
<td>ArA3212</td>
<td>Ivory Coast</td>
<td>1981</td>
<td>(M) Culex guiarti</td>
</tr>
</tbody>
</table>

* 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’ GAGGTTCTTTCACTGCT 3’) and amplified using primers WN240 and WN132 (5’ GAAAAACATCAA-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...
Envelope gene variability of West Nile viruses

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 & Sokley, 1973; Hammam et al., 1965; Hammam & Price, 1966; Mathiot et al., 1990; Odelola & Fabiyi, 1976; Price & O’Leary, 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

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...
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., 1985). SEM and CAR 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 found 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 Aicha Diouf, Annick Leclerc and Mireille Mondo for help in cell culture and virus propagation procedures. We thank Michèle Bouloy and Bhalouli 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


Received 6 March 1997; Accepted 14 May 1997