Geographic distribution and evolution of yellow fever viruses based on direct sequencing of genomic cDNA fragments

Loïc Lepiniec, Lynn Dalgarno, Vu Thi Que Huong, Thomas P. Monath, Jean-Pierre Digoutte and Vincent Deube

1 Institut Pasteur, Department of Virology, 25 rue du Dr Roux, 75724 Paris cedex 15, France,
2 Division of Biochemistry and Molecular Biology, Faculty of Science, The Australian National University, Canberra, Australia 2601, 3 OraVax Inc., 230 Albany Street, Cambridge, Massachusetts 02139, U.S.A. and 4 Institut Pasteur, BP 220, Dakar, Senegal

We have compared the nucleotide sequence of an envelope protein gene fragment encoding amino acids 291 to 406 of 22 yellow fever (YF) virus strains of diverse geographic and host origins isolated over a 63 year time span. The nucleotide fragment of viral RNA was examined by direct sequencing of a PCR product derived from complementary DNA. Alignment with the prototype Asibi strain sequence showed divergence of 0 to 21.5% corresponding to a maximum of 5-2% divergence in the amino acid sequence. Taking 10% nucleotide divergence as a cut-off point, the 22 YF virus strains fell into three topotypes which corresponded to different geographical areas, namely West Africa, Central-East Africa, and South America. Two subgroups were defined in West Africa, a genotypic group circulating in the sylvatic zone of the western part of Africa, from western Ivory Coast–Mali to Senegal, and a group responsible for large outbreaks from eastern Ivory Coast–Burkina Faso to Cameroon. Strains from Central–East Africa showed a low ratio of transition:transversion of about 1 instead of 8 to 10 for other strains, when their nucleotide sequences were compared with those of other African strains. This may reflect a more distant relationship between the former strains and the others. No change was observed in the highly conserved amino acid domain encompassing the TGD sequence, an important determinant of flavivirus tropism and pathogenesis. Our results support earlier observations on the genetic relationships between YF isolates established by T1 oligonucleotide fingerprinting and offer a useful tool for the understanding of YF virus distribution and evolution.

Yellow fever (YF) is an acute disease caused by a virus of the family Flaviviridae, which is transmitted to humans or monkeys by Aedes (Ae.) mosquitoes in Africa and South America, and by the vector Haemagogus in the sylvatic areas of tropical South America (Germain et al., 1982; Monath, 1990). YF virus was twice isolated from ticks (Amblyomma variegatum) in nature in the Central African Republic (Saluzzo et al., 1980). YF virus remains endemic in the equatorial forests of Africa and South America and occasionally emerges in human populations causing severe outbreaks. The first strains of YF virus were isolated from humans in 1927 in Ghana (Asibi strain, parent to vaccine strain 17D) and in Senegal (French viscerotropic strain FV, parent to vaccine French neurotropic virus FNV), respectively (Barrett, 1987). It is considered that the virus originated in Africa, and was introduced into the New World in slave-ships bearing active cases of YF or infected Ae. aegypti.

In Africa, attempts to control outbreaks with emergency immunization has reduced the extent and severity of epidemics in urban areas, however large populations still remain at risk in 33 countries. The last three decades have witnessed considerable African YF activity (WHO, 1989, 1991, 1992). During this period YF has probably affected several hundred thousand people, with a fatality of 20 to 30%. Apart from the massive outbreak in Ethiopia in 1959 to 1962 and very recently in Kenya, all the major outbreaks were in West Africa. These occurred in the vicinity of the endemic area in the dry Savannah, referred to as the ‘emergence zone’ by Germain et al. (1982), or in the moist Savanna. Ae. aegypti is the principal vector in epidemics emerging in the Sahelian Savannah belt of western Africa whereas sylvatic vectors (Ae. furcifer, lutocephalus, africanus) are mostly involved in rural interhuman transmission (Fig. 1).

YF viruses are characterized by a single-stranded positive-sense RNA of 10,862 nucleotides. This encodes a single polyprotein of about 3400 amino acid residues in the gene order 5’ C–prM–E–NS1–NS2A–NS2B–NS3–NS4A–NS4B–NS5 3’ (Rice et al., 1985). The YF virions contain a nucleocapsid C and two surface proteins, a membrane-associated M and an envelope glycoprotein E.
which may or may not be glycosylated (Deubel et al., 1987; Ballinger-Crabtree & Miller, 1990). Based on RNase T1 oligonucleotide fingerprinting of the genomic viral RNA, we have previously classified YF virus isolates into four geographic variants (topotypes) representing western and southern parts of West Africa, Central Africa and South America (Deubel et al., 1986a). However, oligonucleotide fingerprinting of the entire RNA genome is not convenient for the study of a large number of isolates. Its sensitivity to point mutation reduces the interpretation of evolutionary range and is limited to viruses which are less than 5% divergent in their genomic RNA sequence. Moreover, it is not possible to correlate genetic variation with modifications to structural components.

In this study, we have examined a large panel of YF virus isolates including some from older (Ghana 1927, Zaire 1959, Ethiopia 1961, Senegal 1927 and 1965) and recent (Burkina Faso 1983, Nigeria 1986, Mali 1987, Cameroon 1990) epidemics, by direct sequencing of a gene fragment obtained from genomic RNA subjected to reverse transcriptase (RT) and PCR. This methodology was valuable for epidemiological analysis and classification of dengue 2 viruses (Deubel et al., 1993) and was therefore developed to support epidemiological investigations of YF. We have examined the primary nucleotide sequence of an E protein gene fragment of 21 African YF virus isolates, including the two original isolates obtained from humans in Ghana (Asibi) and in Senegal (FV) in 1927. Our sequences have also been compared with the published sequence of South American YF virus strain 1899/81 (Ballinger-Crabtree & Miller, 1990).

Table 1 indicates the location, year of isolation and source of each strain used in this study. All strains dated before 1987 were passaged up to six times in mice or cell culture systems. Since then, all viruses had been propagated in Ae. pseudoscutellaris AP61 cells. For viral RNA preparation, viruses were passaged once in AP61 cells and viral RNA was phenol-extracted from the cytoplasmic lysate of infected cells and ethanol-precipitated as previously described (Deubel et al., 1993).

First strand cDNA was primed and DNA was subsequently amplified using 20-mer oligonucleotide primers chosen to frame an E gene region. The choice was based on the following. (i) Limited sequencing studies performed on the E gene indicated a uniform rate of random nucleotide mutations. (ii) The region encodes a protein domain relevant for viral antigenicity and pathogenicity (Mandl et al., 1989; Lobigs et al., 1990; Mégret et al., 1992). (iii) Amino acid sequence comparison between Asibi and 17D strain E proteins showed two major
Table 1. Strains of yellow fever analysed in this study

<table>
<thead>
<tr>
<th>Code</th>
<th>Origin</th>
<th>Source</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asibi*</td>
<td>Ghana</td>
<td>Human</td>
<td>1927</td>
</tr>
<tr>
<td>FV*</td>
<td>Senegal</td>
<td>Human</td>
<td>1927</td>
</tr>
<tr>
<td>4144 LSF-4*</td>
<td>Zaire</td>
<td>Human</td>
<td>1959</td>
</tr>
<tr>
<td>2777</td>
<td>Ethiopia</td>
<td>Human</td>
<td>1961</td>
</tr>
<tr>
<td>HD 1279*</td>
<td>Senegal</td>
<td>Human</td>
<td>1965</td>
</tr>
<tr>
<td>AMT/7</td>
<td>Ivory Coast</td>
<td>Human</td>
<td>1973</td>
</tr>
<tr>
<td>ArD 24553</td>
<td>Senegal</td>
<td>Aedes furcifer</td>
<td>1976</td>
</tr>
<tr>
<td>ArD 25112</td>
<td>Senegal</td>
<td>Aedes luteocephalus</td>
<td>1977</td>
</tr>
<tr>
<td>ArB 8883</td>
<td>CAR</td>
<td>Aedes africanus</td>
<td>1977</td>
</tr>
<tr>
<td>WBT 1927</td>
<td>CAR</td>
<td>Amblyomma variegatum</td>
<td>1978</td>
</tr>
<tr>
<td>ArD 26923</td>
<td>Senegal</td>
<td>Erythrocebus patas</td>
<td>1978</td>
</tr>
<tr>
<td>ArD 27797†</td>
<td>Senegal</td>
<td>Aedes aegypti</td>
<td>1979</td>
</tr>
<tr>
<td>YF 1899/81*</td>
<td>Peru</td>
<td>Human</td>
<td>1981</td>
</tr>
<tr>
<td>Ha 722*</td>
<td>Burkina Faso</td>
<td>Human</td>
<td>1983</td>
</tr>
<tr>
<td>HD 38559*</td>
<td>Burkina Faso</td>
<td>Human</td>
<td>1983</td>
</tr>
<tr>
<td>HB1782</td>
<td>CAR</td>
<td>Human</td>
<td>1985</td>
</tr>
<tr>
<td>BA 55*</td>
<td>Nigeria</td>
<td>Human</td>
<td>1986</td>
</tr>
<tr>
<td>ArA 20581*</td>
<td>Mali</td>
<td>Aedes furcifer</td>
<td>1987</td>
</tr>
<tr>
<td>ArD 76320</td>
<td>Senegal</td>
<td>Aedes furcifer</td>
<td>1990</td>
</tr>
<tr>
<td>HD 78359*</td>
<td>Cameroon</td>
<td>Human</td>
<td>1990</td>
</tr>
<tr>
<td>ArB 28153</td>
<td>CAR</td>
<td>Aedes gr. tarsalis</td>
<td>1991</td>
</tr>
<tr>
<td>ArA 28105</td>
<td>Ivory Coast</td>
<td>Aedes furcifer</td>
<td>1991</td>
</tr>
</tbody>
</table>

* These strains were associated with an epidemic of YF.
† This strain was isolated 1 year after the epidemic in The Gambia on the border of the country, and was associated with human cases.
‡ YF 1899/81 strain was sequenced previously (Ballinger-Grabert & Miller, 1996).

Changes in the corresponding region. Two amino acid changes, threonine to arginine at position 380 related to RGD sequence, and proline to histidine at position 390, may alter flavivirus attachment or entry (Hahn et al., 1987). (iv) The gene fragment contained an N-glycosylation site of the form asparagine-proline-threonine at amino acid position 309 which could serve as a carbohydrate attachment site (Schlesinger et al., 1983; Rice et al., 1985; Deubel et al., 1987).

Two primer sequences were selected among several DNA probes complementary to 17D strain YF viral RNA (Rice et al., 1985) for cross-reactive hybridization (Kerschner et al., 1986) against the different YF RNAs (data not shown). The oligonucleotide used for cDNA priming on RNA from West African YF strains was complementary to 5'-AUGGGGAUGGCTGCUUGGG-GCAGC3', which corresponds to nucleotides 1237 to 1256 in the 17D strain YF virus E gene. For cDNA priming of the East and Central African strains of YF virus we used the oligonucleotide 5'-TCTTGGTAGGAGTGATCTAGT-GCAG3' complementary to the E protein gene from nucleotides 1433 to 1452. cDNA was obtained by reverse transcription using 0.03 pmol of infected cell RNA, 5 pmol of YF virus-specific primer and a mixture made of RT buffer, RNase inhibitor, the four deoxynucleoside triphosphates and avian myeloblastosis virus RT, as previously described (Deubel et al., 1990, 1993). cDNAs were denatured at 95 °C for 5 min and then subjected to a 30-cycle amplification (denaturation at 95 °C for 15 s, annealing at 55 °C for 60 s and elongation at 72 °C for 60 s) by PCR using 50 pmol of the same antigenomic-sense primer as used for the RT reaction, 50 pmol of a second primer of genomic sense 5'-CAAGCTGCATG-GGGGCACGC3' (nucleotides 822 to 841 in the E protein gene) and a mixture containing PCR buffer, the four deoxynucleoside triphosphates and Taq polymerase (Deubel et al., 1990, 1993). Subsequently, the DNA was purified from 1% low melting-point agarose. Sequencing reactions were carried out on the dsDNA by the dideoxynucleotide chain termination method using either the genomic-sense, YF virus-specific, nucleotides 822 to 841 primer or the antigenomic-sense nucleotides 1237 to 1256 or 1433 to 1452 primers and T7 polymerase (Sequenase) in a standard method (Deubel et al., 1993). Sequencing was performed on DNA products obtained in duplicate from RNA samples.

Sequences of nucleotides 871 to 1218 in the E protein gene fragment from 22 YF virus isolates are shown in Fig. 2. Asibi and FV strains differed by only one nucleotide at position 1119 confirming that the two viruses are closely related and were presumably responsible for all the outbreaks occurring along the West African coast in 1927 (Deubel et al., 1986b; Jennings et al., 1993). The published sequence of the Asibi strain (Hahn et al., 1987) shows two additional nucleotide changes at positions 992 and 1169 in the E protein gene that were not observed in the strain analysed here. This heterogeneity may be due to variation in the passage history of the Asibi strain or to clonal differences (Hahn et al., 1987). Since our RNAs were not derived from plaque-purified viruses, the sequences presented in Fig. 2 may reflect those of the major population in the viral mixture. Despite the temporal separation of 14 years and the different sources of virus isolates in two separate regions of YF endemicity (Germain et al., 1982), five strains from Senegal (1976 to 1990) showed the same nucleotide sequence in the gene fragment analysed and four strains from Central African Republic (1977 to 1991) showed less than 1% nucleotide change, thus confirming the genetic stability of the YF virus in each ecological zone previously noted in T1 oligonucleotide maps (Deubel et al., 1985, 1986a). Nucleotide changes are scattered throughout the E protein gene fragment with a maximum divergence of about 21-5% between the East African strains and the single South American YF virus strain examined. It is intriguing that when compared to the Asibi strain, transitions in West African and South American strains occur eight to 10 times more frequently than transversions, whereas East and Central African virus strain transversions are as common as transitions (Fig. 2). This may reflect cumulative transversions that have become fixed over time and are less subject to back-mutation during divergent strain evol-
Fig. 2. Sequence of 348 nucleotides in the E protein gene of 22 YF virus strains from Africa and South America. The countries and year of isolation are indicated. A strict sequence identity was observed among five Senegalese strains isolated between 1976 and 1990 and among three Central African Republic strains isolated between 1977 and 1985; therefore, only one representative sequence of each of these strains is shown. The identities with the Asibi prototype strain are indicated by dots. The nucleotide positions in the E gene sequence are numbered. Nucleotide sequence mismatches observed in our Asibi sequence when compared with that previously published by Hahn et al. (1987) are indicated by an arrow. Nucleotide transversions occurring in topotypes when compared with the Asibi sequence are indicated by asterisks.

Fig. 3. Amino acid sequences deduced from the 348 nucleotides used to determine the relationships among YF virus isolates. Amino acids 291 to 406 characterized in the YF E protein sequence are presented. The dots indicate amino acids identical to those of the Asibi prototype strain. The TGD sequence is overlined. The putative N-glycosylated site is indicated by an asterisk.

The deduced sequences of the E protein amino acids 291 to 406 for the 22 YF virus strains are aligned with the corresponding region of the Asibi strain in Fig. 3. No amino acid change occurred among the West African
strains between 1927 and 1991. Central African strains exhibited only one conservative amino acid difference, whereas the East African strains showed two (Zaire) or three (Ethiopia) amino acid changes. An alanine was substituted for a proline at position 333 in these two strains. A higher amino acid divergence (5-2%) was observed between Asibi and strain 1899/81 from Peru. The amino acid substitutions are not uniformly distributed along the 116 amino acid fragment. A strongly hydrophilic area of similarity among flaviviruses starts at proline 369 (Ballinger-Crabtree & Miller, 1990). It has been speculated that non-conservative changes occurring in this domain encompassing the TGD sequence during selection in tissue culture may alter receptor affinities and flavivirus pathogenicity (Hahn et al., 1987; Lobigs et al., 1990). No amino acid differences were observed in the region surrounding the TGD set of amino acids despite the isolation of viruses from different sources. This suggests that strict conservation of the structure in this E domain is required for dual infection of vertebrate hosts and arthropod vectors. On the other hand, the significance of none of the minor amino acid changes observed in the sequence of the gene fragment can be interpreted without further biological data.

The relationships of the YF virus strains, based on the similarity of the E protein gene fragment, are shown in a dendrogram in Fig. 4. The scores recorded from the fast-approximate method of nucleotide sequence comparison of Wilbur & Lipman (1983) were used to construct the dendrogram using the UPGMA cluster analysis method of Sneath & Sokal (1973). The CLUSTAL program established a genetic relationship among YF strains by assigning an equivalent statistical weight to each base substitution (Higgins & Sharp, 1988). Although the resolution was reduced when the rate of variation increased, the genetic relationship calculated on the abscissa of the dendrogram remained accurate. The degree of relatedness agrees with our previous classification based on T1 oligonucleotide mapping and on viral protein profiles (Deubel et al., 1985, 1986a, b, 1987). The African strains can be grouped into two main clusters, a western African cluster and a central–eastern cluster including strains from Ethiopia, Zaire and the Central African Republic.

In western Africa, strains are divided into two subgroups, all having less than 7% divergence from the Asibi prototype strain (Fig. 4). The first subgroup corresponds to the historical strains Asibi and FV, which were responsible for the outbreaks along the West African coast in 1927, and to the endemo-epidemic Senegalese strains, isolated over the succeeding years to the present. For the Senegalese strains, less than 3.5%
divergence occurred in the nucleotide sequence of the gene fragment analysed over a 63 year span. This is up to 10 times lower than the value expected when compared with the rates of evolution of the related dengue flavivirus (Rico-Hesse, 1990; Chu et al., 1989; Deubel et al., 1993).

In West Africa, endemic foci are found predominantly in the rainforest and savannah. The regions bordering the rainforest are zones of emergence where intermediate epidemics can occur (Fig. 1). The viruses stay in endemic foci and emerge when ecological conditions are favourable for human transmission. This is the case in Senegal where the sylvatic circulation of YF virus in eastern Senegal has been followed. Epidemics in Djourbel in 1965 and in The Gambia in 1978 (associated strain ArD27797) were associated with an increase in sylvatic circulation of virus and eventual spill-over to interhuman transmission by Ae. aegypti. The same events occurred in the western part of the Ivory Coast and in Mali in 1987. Viruses from the Ivory Coast (ArA 28105) and Mali (ArA 20581) are closely related (≥ 95.4% similarity) to those circulating in Senegal (Fig. 4). The second subgroup in West Africa includes the epidemic strains isolated during explosive outbreaks in Burkina Faso (1983), in Nigeria (1986) and in Cameroon (1990). The strain similarity indicates that the recent epidemics in Cameroon represent an extension of YF virus activity in Nigeria (WHO, 1991, 1992). Numerous YF cases were observed in Ghana in 1983, and a few cases were reported in Niger, near the southern border of Nigeria (WHO, 1989). No virus strains from these outbreaks were available for genetic study but it is reasonable to assume that those outbreaks were due to the same virus subgroup. A single strain isolated in the Ivory Coast in 1973 (AMT 7) was unique when compared with the strain isolated 18 years later, in the Ivory Coast in 1991, and in other West African countries (up to 91% divergence). Nevertheless, this strain showed 94.3% similarity with the strains of the second subgroup. If we consider 10% of nucleotide sequence divergence to be an arbitrary cut-off for grouping the strains as belonging to a genetic topotype, it is conceivable that the present Western African variants have the same origin (Deubel et al., 1986b). The subgroups recognized in West Africa in recent times may have evolved separately since the circulation of virus by the urban vector Ae. aegypti was reduced due to the massive campaign of vaccination which started about 50 years ago and to the reduction of coastal traffic by sailing ships, which seems to have been responsible for the movement of YF to South America.

Intense sylvatic circulation in Central and East Africa has been recorded but human cases are very rare and occur only at low level or in infrequent epidemics. It is not clear whether the lack of epidemics in Central Africa is due to ecological factors limiting transmission or to a less virulent strain of virus. A large epidemic developed in southern Sudan and in Ethiopia in 1959 to 1962. No evidence of previous circulation of the YF virus could be found in Ethiopia and the outbreak was probably the result of introduction of the virus from afar rather than an emergence from a local, enzootic reservoir (Sérié et al., 1968; Brès et al., 1983). The genetic characterization indicates that the source of the epidemic strain was elsewhere in East Africa, as might be expected from the geographic proximity of Uganda, Sudan and Kenya where YF virus sylvatic circulation has been observed (Sérié et al., 1968; McCrae & Kiry, 1982). The virus isolated in 1961 during the Ethiopian epidemic shared 98.6% similarity with a virus isolated from a human case in Zaire in 1959 where YF is mainly endemic with sporadic cases. Therefore, these strains are closer to each other than to the endemic strains in the more geographically distant Central African Republic (94.5 to 95.4% similarity).

The South American strain constitutes the third topotype, and shows 14 to 16% divergence in nucleotide sequence from West African strains and about 21.5% divergence from Central–East African strains. Nevertheless, the amino acid diversity between South American and African strains was higher than that observed between the two African species. This difference may be related to the antigenic variability observed between South American and African YF viruses (Clarke, 1960). These variations may reflect different ecological selection pressures on viral strains between the two continents.

Phylogenetic analysis of a 348 nucleotide sequence revealed broader genetic relationships between different topotypes established by TI oligonucleotide mapping and provided quantitative data which were easier to interpret. Furthermore, our study located differences that were reliable for determining evolutionary relationships. Knowledge of the distribution of YF is important in establishing accurate research and prevention programmes. This distribution fluctuates according to the development of epidemics. Our study gives new insights into the genetic relationships between YF virus isolates from various geographic areas in Africa. It is now possible to define epidemic and sylvatic foci and to connect epidemics by following the virus circulation using simple molecular analysis tools. This is possible first because variability and genetic evolution of the YF virus in one area seem very low and therefore we are able to classify viruses in subgroups which correspond to the ecological zones of virus propagation. We also found a particularity in the genomic structure of the Central–East African strains regarding the unusual number of transformations when compared with other YF strains. It would be interesting to investigate how these strains have diverged genetically from the West African strains and if
these changes have an implication in the level of replication and transmission in the vector–host cycle.

The authors thank the entomologists of ORSTOM, Drs R. Cordemier, M. Cornet and M. Germain who undertook the sample capture in the field and to Drs A. Georges, N. Karabatsos, R. E. Shope, J. Thonon, J. Vincent and H. Zeller for providing the YF virus strains used in this study. We thank M. T. Drolet for excellent technical expertise. We are grateful to Dr D. G. Huggins for the free use of the CLUSTAL V software. This research was supported by grant no. 88-230 awarded by the Direction des Recherches Etudes et Techniques (Paris).

References


(Received 28 June 1993; Accepted 6 October 1993)