

Phylogenetic and Evolutionary Relationships among Yellow Fever Virus Isolates in Africa

JOHN-PAUL MUTEBI, HEIMAN WANG, LI LI, JULIET E. BRYANT, AND ALAN D. T. BARRETT*

Center for Tropical Diseases, Department of Pathology, University of Texas Medical Branch, Galveston, Texas 77555

Received 15 December 2000/Accepted 8 May 2001

Previous studies with a limited number of strains have indicated that there are two genotypes of yellow fever (YF) virus in Africa, one in west Africa and the other in east and central Africa. We have examined the prM/M and a portion of the E protein for a panel of 38 wild strains of YF virus from Africa representing different countries and times of isolation. Examination of the strains revealed a more complex genetic relationship than previously reported. Overall, nucleotide substitutions varied from 0 to 25.8% and amino acid substitutions varied from 0 to 9.1%. Phylogenetic analysis using parsimony and neighbor-joining algorithms identified five distinct genotypes: central/east Africa, east Africa, Angola, west Africa I, and west Africa II. Extensive variation within genotypes was observed. Members of west African genotype II and central/east African genotype differed by 2.8% or less, while west Africa genotype I varied up to 6.8% at the nucleotide level. We speculate that the former two genotypes exist in enzootic transmission cycles, while the latter is genetically more heterogeneous due to regular human epidemics. The nucleotide sequence of the Angola genotype diverged from the others by 15.7 to 23.0% but only 0.4 to 5.6% at the amino acid level, suggesting that this genotype most likely diverged from a progenitor YF virus in east/central Africa many years ago, prior to the separation of the other east/central African strains analyzed in this study, and has evolved independently. These data demonstrate that there are multiple genotypes of YF virus in Africa and suggest independent evolution of YF virus in different areas of Africa.

Yellow fever (YF) virus causes a viral hemorrhagic fever in humans. The case fatality rate of YF can exceed 50% (17). YF remains a major public health concern in sub-Saharan Africa and tropical South America despite the availability of a safe and effective vaccine. For example, in the period between 1987 and 1991, a total of 18,735 YF cases and 4,522 deaths were reported to the World Health Organization (15). These figures represent the greatest YF activity since 1948 (15). YF is transmitted by the bite of an infected female mosquito, usually *Aedes* species in Africa and *Haemagogus* or *Sabethes* species in South America. YF virus is the prototype member of the genus *Flavivirus*, family *Flaviviridae*. Flaviviruses have a small positive-sense, single-stranded RNA genome. The prototype strain of YF virus is Asibi, and the genome consists of 10,862 nucleotides (6). The genome is arranged into a short 5' noncoding region, a single open reading frame consisting of 10,233 nucleotides that encodes the structural genes C, prM, and E and nonstructural genes NS1, NS2A, NS2B, NS3, NS4A, 2K, NS4B, and NS5, and a 3' noncoding region (1a).

Genetic relationships among wild YF virus strains in Africa are not well understood, primarily because of the limited number of studies on the subject. Although six studies (2, 3, 4, 10, 19, 20) have specifically analyzed genetic and phylogenetic relationships among wild YF strains from Africa, these studies utilized relatively few YF strains, from 3 (20) to 21 (10). This is underrepresentative considering the size of the zone where YF is endemic in Africa and the frequency of YF outbreaks in

this region. Most of these studies (2, 3, 10, 19) showed clear genetic and phylogenetic distinction between east/central and west African YF strains. However, the samples were biased towards west Africa, which is probably attributed to the availability of isolates due to more YF activity in this region than in east and central Africa. For example, Lepiniec et al. (10) analyzed 21 wild strains of YF virus, but only 6 (28.5%) were from east and central Africa.

Previous studies (2, 3, 4, 10, 19, 20) showed that east and central African wild YF strains were closely related genetically and belong to the same genotype. The same studies showed that strains from west Africa were genetically distinct from those in east/central Africa and represented two distinct genotypes. Deubel et al. (4) used oligonucleotide fingerprinting and identified three genetically distinct topotypes of YF virus in Africa, two in west Africa and one in east/central Africa. Subsequently, Lepiniec et al. (10) and Chang et al. (2) analyzed nucleotide sequence variation of the envelope (E) protein gene and described two genotypes, one in east/central Africa and the other in west Africa. In the present study, we examined a panel of 38 spatially and temporally diverse wild YF virus isolates from diverse regions of Africa to elucidate the precise genetic relatedness of wild strains of YF virus in Africa.

MATERIALS AND METHODS

Viruses. Thirty-eight low-passage virus strains isolated between 1927 and 1993 from 13 African countries were used in this study (Table 1). The virus strains were lyophilized stocks obtained from the World Arbovirus Reference Collection at the University of Texas Medical Branch, Galveston, Tex. and Centers for Diseases Control and Prevention, Fort Collins, Colo., except strain 85-82H (13). Twenty-eight (73.7%) were epidemic strains, and 10 (26.3%) were enzootic. Similarly, 30 (78.9%) strains were from human cases, 7 (18.4%) were from mosquitoes, and 1 (2.6%) strain from Uganda (Uganda 72; strain number Z 19039) was from a monkey. The majority of strains (23 [60.5%]) were from west

* Corresponding author. Mailing address: Department of Pathology, University of Texas Medical Branch, Galveston, TX 77555-0609. Phone: (409) 772 6662. Fax: (409) 747 2415. E-mail: abarrett@utmb.edu.

TABLE 1. YF virus strains used

Strain	Origin ^a	Year	Source	Code (reference)	GenBank accession no.
14 FA	Angola	1971	Human	Angola71	AF369669
HD 38564	Burkina Faso	1983	Human	Bfaso83a	AF369670
SH 28580	Burkina Faso	1983	Human	Bfaso83b	AF369671
Ar B 8883	C.A.R.	1977	Mosquito	Car77a (18, 19)	U52392
Ar B 9005	C.A.R.	1977	Mosquito	Car77b (18, 19)	U52395
Ar B 17239	C.A.R.	1980	Mosquito	Car80	AF369672
HB 1504	C.A.R.	1985	Mosquito	Car85	AF369673
Serie 227	Ethiopia	1961	Human	Ethiopia61a	AF369674
Couma	Ethiopia	1961	Human	Ethiopia61b	AF369675
Asibi	Ghana	1927	Human	Ghana27 (6)	
85-82H	Ivory Coast	1982	Human	Ivory Coast82 (13)	U54798
BC 7914	Kenya	1993	Human	Kenya93	AF369676
69056	Nigeria	1946	Human	Nigeria46 (18, 19)	U52403
IB AR 45244	Nigeria	1969	Mosquito	Nigeria69	AF369677
T. Adeoye	Nigeria	1970	Human	Nigeria70a	AF369678
A. Jimo	Nigeria	1970	Human	Nigeria70b	AF369679
M. Adejumo	Nigeria	1970	Human	Nigeria70c	AF369680
A. Adeoye	Nigeria	1970	Human	Nigeria70d	AF369681
H117491	Nigeria	1987	Human	Nigeria87a	AF369682
H117505	Nigeria	1987	Human	Nigeria87c	AF369683
BA 55	Nigeria	1987	Human	Nigeria87b	AF369684
56205	Nigeria	1991	Human	Nigeria91	AF369685
JoseCachatra	Guinea-Bissau	1965	Human	GuinB65	AF369686
FVV	Senegal	1927	Human	Senegal27 (18, 19)	GI 694115
Rendu	Senegal	1953	Human	Senegal53 (20)	U89338
SH 1446	Senegal	1965	Human	Senegal65a	AF369687
SH 1464	Senegal	1965	Human	Senegal65b	AF369688
Dak 1279	Senegal	1965	Human	Senegal65c (18, 19)	U52413
SH 1339	Senegal	1965	Human	Senegal65d	AF369689
Dar Ar 276	Senegal	1965	Mosquito	Senegal65e	AF369690
ArD93388	Senegal	1992	Mosquito	Senegal92	AF369691
M 90-5	Sudan	1940	Human	Sudan40a	AF369692
M 112-4	Sudan	1940	Human	Sudan40b	AF369693
A 709-4-A2	Uganda	1948	Human	Uganda48a	AF369694
MR 896	Uganda	1948	Human	Uganda48b (18, 19)	U52422
SE 7445	Uganda	1964	Human	Uganda64	AF369695
Z 19039	Uganda	1972	Monkey	Uganda72	AF369696
LSF 4	Zaire	1958	Human	Zaire58	AF369697

^a C.A.R., Central African Republic.

Africa. Nine strains were from eastern Africa, five from central Africa, and one (Angola 71; strain 14 FA) was from southern Africa. Each reconstituted virus preparation was passaged once in Vero cell cultures to produce seed virus. The seed virus was used to prepare working stocks by an additional passage in Vero cell culture.

Nucleotide sequencing studies. Methods used to grow virus and to extract viral RNA have been described in detail elsewhere (18). Reverse transcription (RT)-PCR was performed on purified viral RNA using the procedures described by Wang et al. (19). One set of studies involved amplification of a 670-bp DNA fragment for all 38 YF strains using the CAG (CTGTCCAATCTCAGTCC) and the YF7 (AATGCTTCCTTCCCAAAT) primers. This fragment included the 3' 108 nucleotides of the premembrane (prM) protein gene, the entire 225 nucleotides of the membrane (M) protein gene, and the 5' 337 nucleotides of the envelope (E) protein-coding gene. We have previously shown that this region is a representative sample of the YF genome (19). The second set of studies amplified the structural protein genes C, prM/M, and E for the Angola strain of YF virus. PCR products were screened using agarose gels and ethidium bromide staining. The PCR products were then extracted from the gels and either cloned into pGEM (Easy) vectors (Promega) or directly sequenced, depending on the quantity of cDNA recovered. For cloned cDNAs, three clones were sequenced to provide a representative consensus sequence for the strain. Sequencing was done using an ABI automatic sequencer at the University of Texas Medical Branch protein chemistry core facility.

Sequence data analysis. Nucleotide sequences for the YF strains were imported directly and aligned using Vector NTI sequence analysis program (Informax). Percentage similarities and differences were calculated using the MegAlign program (DNASTAR, Lasergene). Phylogenetic analysis of the aligned se-

quences was performed using PAUP (16) and NEIGHBOR, a neighbor-joining program, in the PHYLIP package (5). Parsimony analysis was implemented using the heuristic algorithm. The one-parameter formula was used to generate the distance matrix for neighbor-joining analysis (9). Bootstrap analysis with 1,000 resamplings was used to determine confidence values for groupings within the phylogenetic tree. The tree was rooted using a homologous sequence of dengue-1 virus (GenBank accession no. NC001477). We estimated nucleotide substitution rates by identifying sister sequences that were closely related and isolated at least 5 years apart. The differences in changes depicted in branch lengths separating each sister sequence from the predicted common ancestor were divided by the number of years between the sequences and the number of nucleotides in the sample sequence. Several estimates were compared to provide an estimate mean and standard deviation.

RESULTS

Nucleotide sequence variation among strains of YF virus from Africa. We selected a 670-nucleotide fragment that includes coding regions for three proteins, the premembrane (prM), membrane (M), and envelope (E), which has been shown to be representative of the entire genome of wild-type strains of YF virus (19). Nucleotide sequences of this region were determined for 31 strains of YF virus amplified in this study plus 7 previously published sequences and used to generate a phylogenetic tree (Fig. 1). The phylogenetic tree gen-

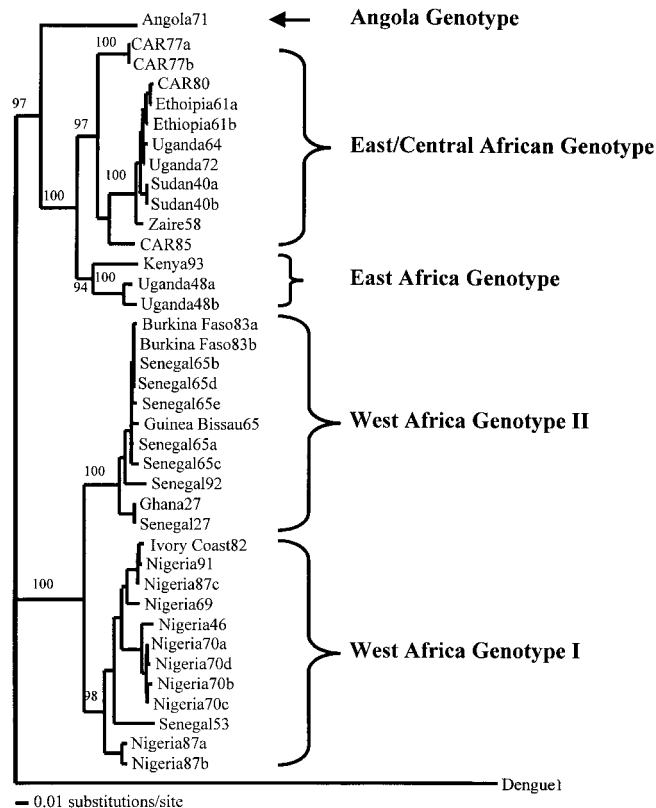


FIG. 1. Neighbor-joining phylogram derived from the nucleotide sequence of the prM/E region of 37 wild YF virus strains from Africa.

erated indicated that wild-type YF virus strains in Africa could be divided into two major lineages, strains in west Africa and those from east/central Africa (Fig. 1). The two lineages were further divided into five clades, two in west Africa and three in east/central Africa, based on phylogenetic relationships (Fig. 1) and variation in nucleotide sequence (Table 2 and Fig. 2). Genotypes were defined as distinct lineages that differed by greater than 9% at the nucleotide sequence level (Table 2). A similar criterion was used by Lepineic et al. (10), Wang et al. (18), and Chang et al. (2) to define YF virus genotypes in Africa and South America.

We named the genotypes according to the region of origin of the strains. West African genotype I included strains from Nigeria, Ivory Coast, and one strain, Rendu (Senegal53), from Senegal. West African genotype II included strains from Ghana, Senegal, Burkina Faso, and Guinea-Bissau (Portu-

guese Guinea). In east and central Africa, the Angola genotype consisted of a single strain, 14 FA (Angola71), from Angola (12). The east/central African genotype included strains from Central African Republic, Ethiopia, Uganda, Sudan, and Democratic Republic of the Congo (formerly Zaire). The east African genotype consisted of three strains, two from Uganda, A709-4-A2 (Uganda 48a) and MR 896 (Uganda48b), and one from Kenya, BC 7914 (Kenya93) (Fig. 1 and 2 and Table 2).

Pairwise comparisons of the nucleotide sequences showed 0 to 25.8% variation (74.8 to 100% identity) among the 38 wild strains of YF virus (Tables 2 and 3). The nucleotide sequence alignment separated the strains into two major lineages, west Africa and east/central Africa, by substitutions at 19 diagnostic nucleotide positions. A nucleotide position was considered diagnostic if all members of one genotype were fixed for a particular nucleotide at that position and the other genotypes were fixed for alternative nucleotides at the same position. Twenty-one diagnostic nucleotide substitutions separated the two genotypes in west Africa, while the three genotypes in east/central Africa were collectively characterized by two diagnostic nucleotide substitutions. Figure 2 shows a portion of the alignment, including the 3' 108 nucleotides of the prM gene, highlighting nucleotide differences among the genotypes. Pairwise comparisons of the east/central Africa genotype showed that the Angola71 genotype was differentiated from the east African genotype (Uganda48a and -48b and Kenya93) and the east/central African genotype at 63 and 56 nucleotide positions, respectively. The east African and the east/central African genotypes were differentiated at 10 nucleotide sites.

Transition-to-transversion ratios were computed for strains representing the different genotypes. All pairwise comparisons were made to the prototype strain, Ghana27 (Asibi). Strains from east and central Africa had transition/transversion ratios close to 1.5:1, whereas those in west Africa were 6 to 14 times greater (Table 4). Specifically, west Africa genotype I had a ratio of 5.0:1 to 7.7:1, while west Africa genotype II had a ratio of 14:1 to 18:1.

To substantiate the extensive nucleotide variation of Angola71, the complete nucleotide sequence of the structural protein genes of this strain was determined and compared with those of representatives of three different genotypes (Ghana27, CAR77b, and Peru81). Peru81 was included in the study for comparison with a South American strain of YF virus. Pairwise comparison of the nucleotide sequences showed ranges greater than 9% (i.e., 13.8 to 19.1%) (Table 5), which was the cutoff value that we used to define genotypes and confirmed the results obtained with the prM/E region.

TABLE 2. Nucleotide (boldface) and amino acid (lightface) sequence variations among genotypes of wild-type YF virus strains from Africa^a

Genotype	% Variation				
	Angola	East Africa	East/central Africa	West Africa I	West Africa II
Angola	—	0.4–0.9	0.9–4.2	4.2–5.1	4.6–5.6
East Africa	16.7–17.4	—	0.0–4.6	3.7–5.6	4.2–6.1
East/Central Africa	15.7–18.5	8.0–11.9	—	4.2–8.6	4.6–9.1
West Africa I	21.1–23.0	19.9–21.7	19.4–25.4	—	0.4–2.3
West Africa II	22.4–22.8	22.4–23.8	21.9–25.8	6.7–11.0	—

^a The variations show ranges of pairwise comparisons among strains in two different genotypes.

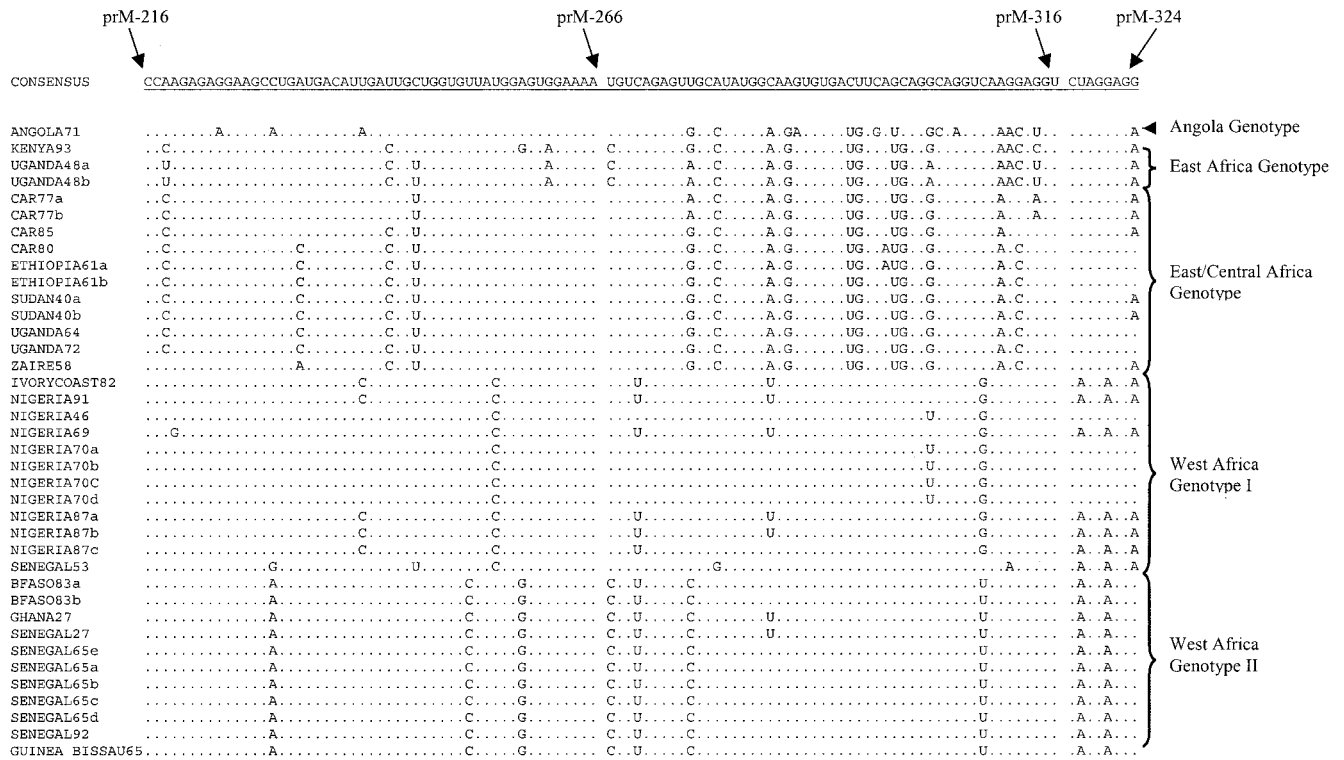


FIG. 2. Nucleotide sequence alignment for the distal 108 nucleotides of the premembrane protein gene of 37 YF viruses from tropical Africa. This alignment highlights nucleotide variations among different genotypes and similarities within genotypes. Dots indicate identity with the consensus sequence.

Nucleotide variation differs between genotypes. Nucleotide variation in west African genotype II was only 2.8% despite the fact that the strains were from four different countries (Senegal, Burkina Faso, Guinea-Bissau, and Ghana) and were isolated up to 65 years apart (1927 to 1992). Similarly, a clade within the east/central African genotype (Fig. 1) consisted of eight strains (CAR80, Ethiopia61a and -61b, Uganda64, Uganda72, Sudan40a and -40b, and Zaire58) from five countries (Central African Republic, Ethiopia, Uganda, Sudan, and Zaire) that were isolated over a period of 40 years. Again, nucleotide variation within this clade was only 2.4%. In contrast, nucleotide variation in west African genotype I (predom-

inantly strains from Nigeria) was up to 6.8% for strains isolated over a 45-year period and was more than twice that observed for the two clades above. Also, four strains isolated in the Central African Republic over an 8-year period (1977 to 1985) and all within the east/central African genotype varied by up to 7.6% (Table 6).

The east Africa genotype, which included strains Kenya93, Uganda48a, and Uganda 48b, was intriguing. The Ugandan strains were isolated in 1948, 45 years before the isolation of strain Kenya93, and differed up to 7.7% at the nucleotide level (Table 3). Second, other strains from Uganda (isolated in 1964 and 1972) were genetically differentiated from the isolates

TABLE 3. Nucleotide and amino acid variation within genotypes of YF virus strains from Africa

Genotype	Distribution	No. of strains	% Nucleotide variation	% Amino acid variation
Angola	Angola	1		
East Africa	Uganda, Kenya	3	1.7–7.7	0.4–1.4
East/central Africa	Central African Republic, Ethiopia, Uganda, Sudan, Zaire	11	0.0–8.3	0.0–4.2
West Africa I	Nigeria, Ivory Coast, Senegal	12	0.1–6.8	0.0–1.8
West Africa II	Ghana, Senegal, Burkina Faso, Guinea-Bissau	10	0.0–2.8	0.0–1.8

TABLE 4. Transition/transversion ratios for 12 representative YF virus strains from Africa^a

Genotype	Strain	Transition/transversion ratio
Angola	Angola71	1.5:1
East/central Africa	Car77a	1.5:1
	Uganda64	1.6:1
East Africa	Ethiopia61b	1.5:1
	Uganda48a	1.5:1
	Kenya93	1.6:1
West Africa I	Nigeria70a	6.3:1
	Nigeria46	5.0:1
	Ivory Coast 82	7.7:1
West Africa II	Senegal65a	14.0:1
	Bfaso83b	15.0:1
	Guinea-Bissau65	18.0:1

^a Ratios were calculated using Ghana27, the prototype strain, as the reference sequence.

from 1948 and differed up to 10.7% at the nucleotide level (Table 6), suggesting that two genotypes of YF virus have been circulating in Uganda. The two other isolates from Uganda (Uganda64 and Uganda72) were isolated from central Uganda, the Bulemezi District and Zika Forest, respectively, approximately 50 miles apart. The geographical origins of Uganda48a and Uganda48b are unclear from the literature. It is well known that the YF epidemic in Kenya in 1993 was in western Kenya, close to the border with Uganda. Uganda48a and Uganda48b are very similar at the nucleotide level (98.3% identity), which was expected because they were isolated in the same epidemic. However, the two strains are 6.7 and 7.7% different, respectively, from the Kenya93 strain. This may be due to the 45 years separating the strains, and the Ugandan strains of 1948 may represent a stage in the evolution of this genotype

High degree of amino acid sequence homology between strains of YF virus from Africa. In comparison to the extensive nucleotide variation, the deduced amino acid sequences for all 38 YF virus strains showed a high degree of sequence homology (91.9 to 100%). The variable amino acid positions did not show genotype-specific differences similar to those defined by the nucleotide analyses above. However, by comparison to a consensus YF sequence of all 38 strains used in this study, amino acid variations at eight amino acid positions segregated all the strains into two geographically distinct groups, west Africa and east/central Africa (Fig. 3). These two major groups

TABLE 5. Nucleotide (boldface) and amino acid (lightface) sequence similarities for representative strains of four genotypes^a

Strain	% Sequence similarity			
	Angola71	Car77b	Ghana27	Peru81
Angola71	—	97.2	95.8	94.2
Car77b	85.2	—	94.9	93.1
Ghana27	81.1	82.4	—	96.5
Peru81	80.9	81.0	86.2	—

^a The pairwise comparisons were made using sequences of the structural protein genes. Sources of nucleotide sequence were as follows: for Angola71, this paper; for Car77b, Wang et al. (19), accession no. U52392; for Ghana27, Hahn et al. (6) and Chang et al. (2), accession no. U23571; for Peru81, Ballinger-Crabtree and Miller (1), accession no. U14458.

TABLE 6. Nucleotide variation within YF virus strains from Africa

Country	Period	No. of yr	% Nucleotide variation	% Amino acid variation
Central African Republic	1977–1985	8	0.0–7.6	0.0–2.7
Senegal	1927–1965	38	0.1–11.2 ^a	0.0–1.4
Nigeria	1946–1991	45	0.1–6.6	0.0–1.8
Uganda	1948–1972	24	0.7–10.7	0.4–1.8

^a Without strain Rendu, 0.1 to 2.8%.

were characterized by four amino acid substitutions (amino acids [aa] 96, 99, 100, and 104) towards the C terminus of the prM protein, one (aa 55) in the M protein, and three (aa 46, 62, and 87) in the N terminus of the E protein (Table 3).

In the prM protein, the YF virus consensus sequence and all west African strains had the amino acid residues Lys, Ser, Ala, and Arg at positions 96, 99, 100, and 104, respectively. The east/central African YF virus strains had the Lys at position 96 substituted with Arg, the Ser at position 99 substituted with Ala, the Ala at position 100 substituted with Val or Met, and the Arg at position 104 substituted with a Lys (Fig. 3). In the M protein, discriminating amino acid substitutions occurred at residues 44 and 55. The consensus YF virus sequence and all west African strains had amino acid Val and Ser at these positions, respectively, whereas most east/central African isolates had amino acid substitutions of Ile and Asn, respectively, at the same positions. In the portion of the E protein analyzed, three amino acid substitutions were observed. One was at position 46, where the consensus YF virus sequence and all west African strains had Glu, while all east/central African strains had Gln at the same position. Similarly, residues 62 and 87 of the consensus YF virus sequence and the west African strains were Asn and Glu, respectively, whereas members of the east/central Africa genotype had Ser and Asp, respectively, in the same positions.

In comparison to the nucleotide variation, west African strains could only be divided into two groups by a single amino acid substitution at position 48 in the membrane protein. One group containing strains from west Africa genotype I (Nigeria and Ivory Coast) had Ala at this position, whereas west Africa genotype II, including strains from Senegal, Ghana, Guinea-Bissau, and Burkina Faso, had Thr at this position. However, Nigeria69 (IBAR45244) had the same amino acid substitution (Thr) at position M-48 as west African genotype II viruses.

The amino acid sequence of the Angola71 strain was very similar to the amino acid sequences of the other east/central African strains despite relatively large differences in the nucleotide sequence. The only variation in amino acid sequence was at position 44 of the M protein, where Angola71 had Ile instead of the Val found in most of the other east/central African isolates. Two strains, Uganda48a (A 709-4-A2) and Uganda48b (MR 896) had amino acid sequences very similar to that of Angola71. Uganda48a had only a single amino acid difference from Angola71 at residue 57 of the M protein, whereas Uganda48b and Angola71 were identical at the amino acid level.

Amino acid sequence variation for the structural proteins of representatives of four different genotypes (Ghana27, CAR77b, Angola71, and Peru81) is summarized in Table 5. As with the

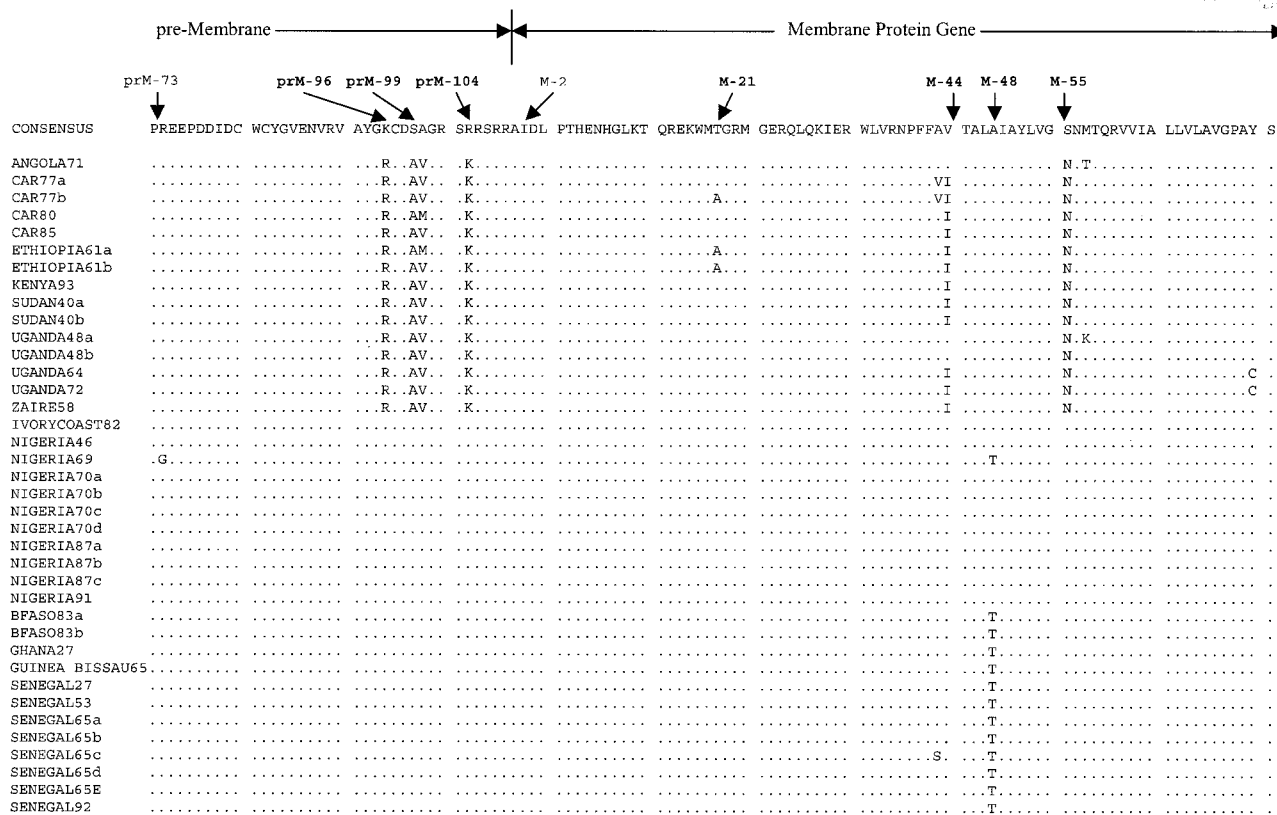


FIG. 3. Amino acid sequence alignment for the prM/E region of 37 YF viruses from Africa. Group-specific amino acid residues are highlighted for each genotype. Dots indicate identity with the consensus sequence.

prM/E region, there was little amino acid variation throughout the structural protein region, and variation ranged from 2.8%, between CAR77b and Angola71, to 6.9%, between CAR77b and Peru81. An alignment of the amino acid sequences for the four strains shows genotype-specific amino acid substitutions (Fig. 4). The most variable region was the 20 amino acids at the carboxy terminus of the C protein. There were significant substitutions in the E protein, at residues 46 (Glu to Gln), 89 (Asp to Gly), 268 (Thr to Glu), 270 (Asp to Trp), and 275 (Lys to Arg) (Fig. 4). The east and central African strains were different from Ghana27 at 25 positions and different from Peru81 at 37 positions. The two east and central African strains CAR77b and Angola77 were different at 21 residues (Fig. 4).

Codon usage differs between genotypes of YF virus. Because of observed genotypic differences in transition/transversion ratios, we investigated codon usage by different genotypes of YF virus in Africa. We compared the number of times a particular codon was used to the expected value within and among YF strains by using the χ^2 test. Expected values for the χ^2 test were calculated assuming that all codons were used at the same frequency. Five YF strains, CAR77b, Nigeria70a, Angola71, Ghana27, and Kenya93, representing the five YF virus genotypes in Africa were selected for this study. Significant bias in codon usage was detected for four amino acids, Leu, Ile, Ala, and Lys (Table 7). However, codon usage for two amino acids (Lys and Ile) separated east/central African strains from west

African strains, while codon usage for Leu separated Angola71 from all other strains (Table 7).

Rates of evolution of different genotypes are very similar. To investigate the nucleotide sequence variation in more detail, we determined the rates of evolution for three genotypes, west Africa I, west Africa II, and east/central Africa, and found that there was no statistically significant difference ($4.58 \times 10^{-4} \pm 7.36 \times 10^{-4}$, $2.344 \times 10^{-4} \pm 1.35 \times 10^{-4}$, and $7.9 \times 10^{-5} \pm 6.03 \times 10^{-4}$ nucleotides per site per year, respectively). We estimated rates of nucleotide substitutions using the times of isolation for the various YF virus strains. For each YF virus genotype, we selected several sister pair sequences isolated less than 7 years apart, with less than 10 nucleotide changes. Small nucleotide changes are better estimators for rates of evolution because multiple substitutions increase with the number of nucleotide changes. The mean for several sister pairs in each genotype was computed and used as the rate of evolution for each genotype.

Contribution of quasi-species to different genotypes. It is well recognized that RNA viruses are continually evolving and consist of quasi-species. To investigate the potential contribution of quasi-species to genetic variation observed within genotypes, an example of each of four genotypes was examined in detail. The 17D vaccine strain was included as a reference. A PCR product for each virus was cloned, and 20 clones of each PCR product were sequenced. The 20 clones were each com-

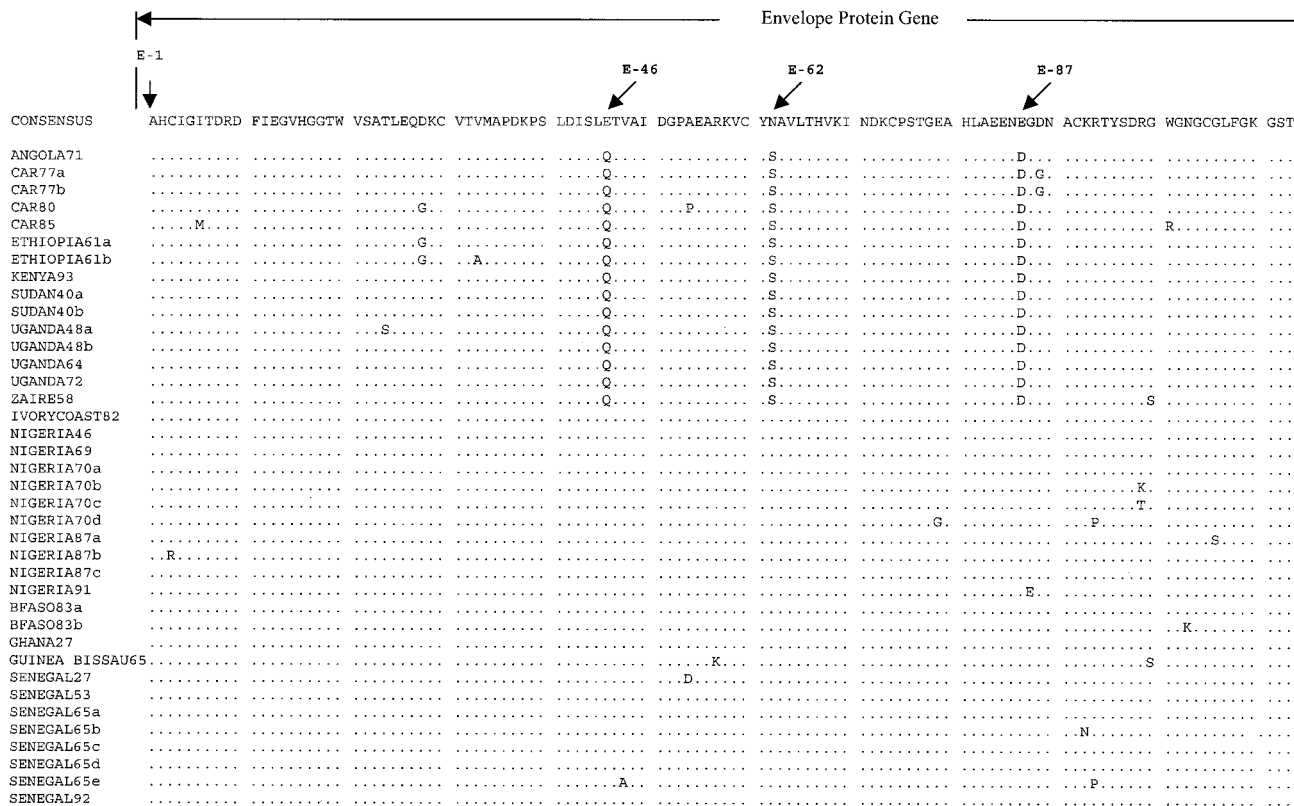


FIG. 3—Continued.

pared to the consensus sequence obtained by direct sequencing of the PCR product, and nucleotide substitutions were identified. The vast majority of the clones were identical in sequence to the consensus sequence, giving confidence in the nucleotide sequence data generated (Table 8). The 17D vaccine and Nigeria69 viruses gave identical results, with 17 clones (85%) identical to the consensus sequence and 3 clones each with a single nucleotide difference. Guinea-Bissau65 was very similar, with only one clone different from the consensus sequence. Thus, the greater nucleotide variation within west Africa genotype I compared to west Africa genotype II could not be explained by quasispecies. The east African viruses Kenya93 and Ethiopia61a revealed greater nucleotide variation within the virus population, with both viruses containing 25 to 40% clones with up to three nucleotide differences from the consensus sequence. Thus, there was greater nucleotide variation within east African viruses compared to west African viruses.

DISCUSSION

The results of the present study support and extend the already established concept that wild-type strains of YF virus in Africa are genetically heterogeneous. We present evidence which suggests that wild-type strains of YF virus in Africa are genetically more diverse than originally reported by Lepiniec et al. (10), Chang et al. (2), and Wang et al. (19). These previous studies used a smaller panel of strains to define two major genetically distinct lineages of YF viruses in Africa, one

in east/central Africa and the other in west Africa. The group in west Africa was further divided into two distinct subgroups (10), but until now it has been thought that only one genotype of YF virus circulated in east and central Africa. In the present study we have described five genotypes, two in west Africa and three in east and central Africa. Although we were able to demonstrate up to 25.8% nucleotide variation between strains, there was a maximum of 9.1% amino acid variation between strains, and there were few amino acid differences between most strains (Fig. 3).

Significantly, nucleotide variation within a given genotype varied greatly among genotypes. Members of west African genotype II and a clade within the east/central African genotype were highly homogeneous ($\leq 2.8\%$ nucleotide variation), whereas west African genotype I was more heterogeneous (up to 6.8% nucleotide variation). The strains in the former two clades are found in sylvatic transmission cycles in enzootic foci that exist year-round in the equatorial rain forest (enzootic zone) and are transmitted predominantly from monkey to monkey by *Aedes africanus* (17). Viruses within these two genotypes have been associated occasionally with epidemics (11) that, presumably, are associated with ecological conditions favoring human transmission. In contrast, there was up to 6.8% nucleotide variation in west Africa genotype I. Strains within this clade were predominantly from moist savannas (zone of emergence), dry savannas, and urban areas (epidemic zone), where transmission is seasonal, involving monkeys and *Aedes furcifer*, *Aedes luteocephalus*, and *Aedes vittatus* (17). Survival

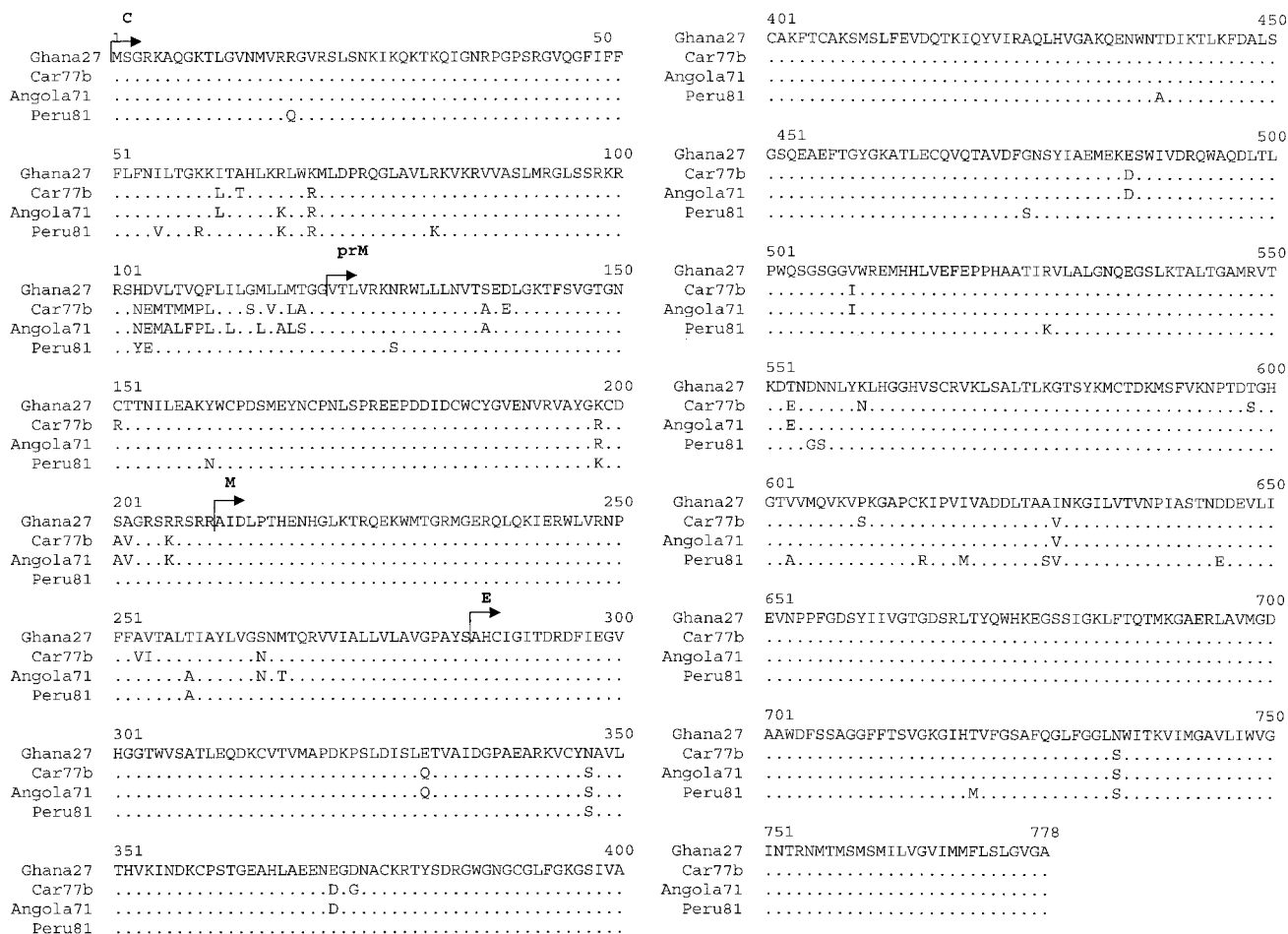


FIG. 4. Amino acid sequence alignment for the structural protein region of four wild-type YF virus strains representing different genotypes.

and continuation of epizootics are ensured by vertical transmission in the mosquitoes. The increased nucleotide variation within this genotype probably reflects adaptation to more diverse conditions in the savannah transmission cycles compared to the stability in the sylvatic environment. Interestingly, the ratio of transitions to transversions was different for the two west African genotypes, providing indirect evidence in support of the two genotypes, existing in different transmission cycles.

The viruses found in the clade including Sudan, Ethiopia, Uganda, Central African Republic, and Zaire have been associated with large epidemics separated by long periods with few human cases. It has been speculated that this phenomenon is due to introduction of viruses from other areas. The phylogenetic analysis supports the hypothesis of continual enzootic activity in these countries with occasional epidemics that are presumably associated with favorable ecological conditions for human transmission. The results also indicate that the recent outbreak in Kenya in 1993 involved viruses genetically related to those isolated in Uganda in 1948 rather than viruses found in the east/central African genotype. Since Angola71 is also an independent lineage, we speculate that multiple independently evolving lineages of YF virus exist in east/central Africa in sylvatic transmission cycles and these viruses only rarely come into contact with humans. In comparison, viruses within west

African genotype I include strains often associated with epidemics, most notably from Nigeria. The frequent human epidemics appear to be associated with extensive genetic variation within this genotype.

Strain14 FA, isolated in Angola in 1971, was well differentiated from all the other strains identified to date at the nucleotide sequence level and was clearly a different lineage phylogenetically. However, the amino acid sequence of Angola71 was remarkably similar to those of other strains in east and central Africa, which suggests that the Angola71 strain probably evolved from a progenitor east/central African virus. Virological studies following the isolation of Angola71 showed that it was antigenically very similar to the Asibi (Ghana27) strain (12), suggesting a close relationship between west African YF virus strains and Angola71. This is consistent with the amino acid identity described in this paper. Moreover, the YF epidemic in Angola in 1971 was the first in Angola for almost 100 years and followed extensive YF activity in several west African countries (Cameroon, Equatorial Guinea, Ghana, Nigeria, and Togo) in 1970. Pinto and Filipe (12) speculated that the Angola outbreak was due either to YF virus moving south from west Africa or to a sylvatic strain. Our results clearly establish a strong genetic relationship between Angola71 and east/central Africa YF virus strains and suggest that the An-

TABLE 7. Codon frequencies for four amino acids representative of five genotypes of wild-type YF virus in Africa^a

Amino acid	Codon	No. of occurrences					Total χ^2
		CAR77b	Nigeria70a	Angola71	Ghana27	Kenya93	
Leu	UUA	0	2	0	0	0	46.2*
	UUG	2	4	1	4	2	
	CUU	0	2	1	1	1	
	CUC	4	2	1	3	3	
	CUA	3	1	8	4	3	
	CUG	6	4	4	3	6	
	χ^2	11	3	18.2*	5.4	8.6	
Ile	AUU	6	7	3	10	6	26.4*
	AUC	2	3	3	1	2	
	AUA	4	1	5	0	4	
	χ^2	2	5.1	0.73	16.57*	2	
Gln	CAA	4	5	1	5	4	13.99*
	CAG	2	0	5	0	2	
	χ^2	0.66	5*	2.7	5*	0.66	
Thr	ACU	3	5	6	7	6	23.45*
	ACC	3	4	2	3	0	
	ACA	7	2	4	2	3	
	ACG	0	2	2	2	4	
	χ^2	7.63	2.1	3.13	4.85	5.77	

^a Statistically significant differences in codon usage were detected at these four amino acids using the χ^2 test. *, significant χ^2 value at the 0.05 level.

golan outbreak in 1971 was due to a sylvatic strain originating in east or central Africa. However, the extensive nucleotide differences between Angola71 and east/central African strains (15.7 to 18.5%) indicate that the progenitor to Angola71 diverged from central/east African strains many years ago. Unfortunately, we have been unable to locate any additional Angolan strains for inclusion in this study.

Earlier work by Deubel et al. (4) and Lepiniec et al. (10) described two distinct YF virus genotypes in west Africa. They observed that what we term west African genotype II was circulating in the area between western Ivory Coast, Mali, and Senegal, while what we term west African genotype I was circulating in the area from eastern Ivory Coast north to Burkina Faso and east to Nigeria and Cameroon (10). Our results confirm the same pattern of distribution of these genotypes, providing more support for the already reported genetic relationships within this region. However, one strain, Rendu, isolated in Senegal in 1953, was genetically distinct from other strains from Senegal in particular and other strains in west African genotype II in general. Studies by Wang et al. (20) showed that the same strain, Rendu, differed significantly from the other strains from Senegal antigenically and at the nucleotide level.

They suggested that Rendu belonged to a different genotype. Our data confirm the observations of Wang et al. (20) and show that Rendu belongs to west African genotype I (the Nigeria and Ivory Coast genotype).

We evaluated nucleotide sequence variation among strains from Central Africa Republic, Senegal, Nigeria, and Uganda (Table 6). These countries were selected because we had data for four or more strains of YF virus from each country. Our data show that nucleotide variation of the strains in Nigeria and Central African Republic was below 9% (i.e., below the level of variation between genotypes). This indicated that the same YF virus genotype was circulating in Nigeria throughout the 45 years between isolation of the earliest and most recently examined strains. Likewise, one genotype was circulating in the Central African Republic. In Senegal, variation of up to 11.2% indicated that more than one genotype was present. However, Senegal53 (Rendu) is considered an imported strain (20). Thus, when Senegal53 is excluded from the analysis, nucleotide variation range was 0.1 to 2.8%, showing that one genotype circulates in Senegal (Table 6). Similarly, nucleotide variation of up to 10.7% in Uganda indicates more than one genotype. We have shown that the strains isolated in 1948 were geneti-

TABLE 8. Nucleotide differences between 20 cDNA clones containing 670 nucleotides from the prM/E region and a consensus sequence determined by direct sequencing of a PCR product

Genotype	Strain	No. of clones with 0, 1, 2, or 3 nucleotide differences compared to consensus sequence				% Nucleotide variation
		0	1	2	3	
West Africa I	17D Vaccine	17	3			0.15
	Nigeria69	17	3			0.15
West Africa II	Guinea-Bissau65	19	1			0.15
East Africa	Kenya93	15	2	2	1	0.45
East/central Africa	Ethiopia61a	12	6	1	1	0.45

cally distinct from those isolated in 1964 and 1972, indicating that at least two YF virus genotypes circulate in Uganda. The above provides evidence that the geographical distribution of strains of YF virus is restricted in Africa and that it evolves very slowly.

Our results show significant bias in codon usage by different genotypes of YF virus in Africa. Differences in codon usage between east/central African genotypes and west African genotypes suggest variations in the enzootic-endemic cycles between these regions. Codon choice has been associated with AT or GC content (7), but the GC content of all the strains that we analyzed in this study was close to 50%, suggesting that it was not a significant determinant of the observed codon bias. Hooper and Berg (7) suggested that there is positive selection on codons that are translated more efficiently, either faster or more accurately. Since YF virus uses host cell macromolecular machinery for replication and protein synthesis, the observed codon usage bias may be attributed to the host. Differences in codon usage bias among the YF genotypes may then be attributed to regional differences among vector species and especially in regions where human YF epidemics are frequent. In such cases, the YF viruses adapt to peridomestic and urban vector mosquito species that may have varying codon usage strategies. We generally assume that the jungle cycle is more stable because it usually involves a single mosquito vector, *A. africanus*, and several species of Old World monkeys. However, genetic variations among regional populations of the jungle vector and vertebrate hosts may also result in the observed codon usage bias among the YF virus genotypes.

Overall, the genetic relationships between strains of YF virus in Africa were more complex than previously described. We have demonstrated that multiple genotypes of YF virus exist in Africa. The results are consistent with the hypothesis that YF virus is undergoing independent evolution in different areas in Africa. Although we have identified clear genotypic differences between strains of YF virus, the phenotypic differences of these viruses remain to be elucidated.

ACKNOWLEDGMENTS

We thank Bob Tesh, Bob Shope, and John Roehrig for the virus strains analyzed in this study and Scott Weaver for help during data analysis.

This work was supported in part by grant AI10986 and by the 2000–2001 Colin Powell Minority Postdoctoral Fellowship in Tropical Disease Research to J.-P.M.

REFERENCES

1. Ballinger-Crabtree, M. E., and B. R. Miller. 1990. Partial nucleotide sequence of South American yellow fever virus strain 1899/81: structural proteins and NS1. *J. Gen. Virol.* **71**:2115–2121.
- 1a. Chambers, T. J., C. S. Hahn, R. Galler, and C. M. Rice. 1990. Flavivirus genome organization, expression, and replication. *Annu. Rev. Microbiol.* **44**:649–688.
2. Chang, G.-J. J., C. B. Cropp, R. M. Kinney, D. W. Trent, and D. J. Gubler. 1995. Nucleotide sequence variation of the envelope protein gene identifies two distinct genotypes of yellow fever virus. *J. Virol.* **69**:5773–5780.
3. Deubel, V., J. P. Paillez, M. Cornet, J. J. Schlesinger, M. Diop, A. Diop, J.-P. Digoutte, and M. Girard. 1985. Homogeneity among Senegalese strains of yellow fever virus. *Am. J. Trop. Med. Hyg.* **34**:976–983.
4. Deubel, V., J.-P. Digoutte, T. P. Monath, and M. Girard. 1986. Genetic heterogeneity of yellow fever virus strains from Africa and the Americas. *J. Gen. Virol.* **67**:209–213.
5. Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5. Department of Genetics, University of Washington, Seattle, Wash.
6. Hahn, C. S., J. M. Dalrymple, J. H. Strauss, and C. M. Rice. 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.
7. Hooper, S. D., and O. G. Berg. 2000. Gradients in nucleotide and codon usage along *Escherichia coli* genes. *Nucleic Acids Res.* **28**:3517–3523.
8. Jennings, A. D., J. E. Whitby, P. D. Minor, and A. D. T. Barrett. 1993. Comparison of the nucleotide and deduced amino acid sequences of the envelope protein genes of the wild-type French viscerotropic strain of yellow fever virus and the live vaccine strain, French neurotropic vaccine, derived from it. *Virology* **192**:692–695.
9. Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules, p. 21–132. *In* H. N. Munro (ed.), *Mammalian protein metabolism*. Academic Press, New York, N.Y.
10. Lepiniec, L., L. Dalgarno, V. T. Q. Houg, T. P. Monath, J.-P., Digoutte, and V. Deubel. 1994. Geographical distribution and evolution of yellow fever viruses based on direct sequencing of genomic cDNA fragments. *J. Gen. Virol.* **75**:415–423.
11. Monath, T. P. 1989. Yellow fever, p. 139–231. *In* T. P. Monath (ed.), *The arboviruses: epidemiology and ecology*, vol. 5. CRC Press, Boca Raton, Fla.
12. Pinto, M. R., and A. R. Filipe. 1973. Arbovirus studies in Luanda, Angola. *Bull. WHO* **49**:31–35.
13. Pisano, M. R., J. Nicoli, and H. Tolou. 1997. Homogeneity of yellow fever virus strains isolated during an epidemic and a post-epidemic period in West Africa. *Virus Genes* **14**:225–234.
14. Rice, C. M., E. M. Lenches, S. R. Eddy, S. J. Shin, R. L. Sheets, and J. H. Strauss. 1985. Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. *Science* **229**:726–733.
15. Robertson, S. E., B. P. Hull, O. Tomori, O., Bele, J. W. LeDuc, and K. Esteves. 1996. Yellow fever: A decade of reemergence. *JAMA* **276**:1157–1162.
16. Swofford, D. L. 1991. PAUP: Phylogenetic Analysis Using Parsimony, version 3.0. Illinois Natural History Survey, Champaign, Ill.
17. Tomori, O. 1999. Impact of yellow fever on the developing world. *Adv. Virus Res.* **53**:5–34.
18. Wang, E., K. D. Ryman, A. D. Jennings, D. J. Wood, F. Taffs, P. D. Minor, P. G. Saunders, and A. D. T. Barrett. 1995. Comparison of the genomes of the wild-type French viscerotropic strain of yellow fever virus with its vaccine derivative French neurotropic vaccine. *J. Gen. Virol.* **76**:2749–2755.
19. Wang, E., S. C. Weaver, R. E. Shope, R. B. Tesh, D. M. Watts, and A. D. T. Barrett. 1996. Genetic variation in yellow fever virus: Duplication in 3' non-coding region of strains from Africa. *Virology* **225**:274–281.
20. Wang, H., A. D. Jennings, K. D. Ryman, C. M. Late, E. Wang, H. Ni, P. D. Minor, and A. D. T. Barrett. 1997. Genetic variation among strains of wild-type yellow fever virus from Senegal. *J. Gen. Virol.* **78**:1349–1352.