Molecular epidemiology of type 1 polioviruses from Africa

Claudia Chezzi, Nigel K. Blackburn and Barry D. Schoub

National Institute for Virology and Department of Virology, University of the Witwatersrand Medical School, Johannesburg, South Africa

The genetic relationships between type 1 polioviruses circulating in sub-Saharan Africa during the past decade have been investigated by partial genomic sequencing across the VP1/2A region of the polioviral genome. Sequencing templates were generated by single-step reverse transcription PCR amplification of the viral RNA using poliovirus-specific primers. Seven poliovirus genotypes, circulating in different geographical regions during different periods, were identified. Considerable genetic variation was exhibited by strains within several of these genotypes, indicative of sustained endemic transmission within individual countries. Two genotypes appear to be circulating in Africa at present; one major genotype, which has been in circulation since at least 1980, covers a wide geographical region and includes countries in western, central and southern Africa. Within this genotype are several smaller clusters, possibly representing strains in the process of evolving into new genotypes. The second genotype presently in circulation has been found only in Tanzania and Zambia to date, associated with a relatively small number of cases. Imported genotypes, introduced from the Middle East and the Indian subcontinent, have also recently been in circulation in eastern and central Africa. In South Africa, three genotypes, one unique to the country and the others imported from west Africa and the Middle East, co-circulated endemically between 1980 and 1985. A fourth genotype, introduced from countries to the north, displaced the endemic strains and continued to circulate until 1989. This study has generated a meaningful overview of the endemic circulation and regional transmission of type 1 polioviruses throughout sub-Saharan Africa.

Introduction

Poliovirus-induced paralytic poliomyelitis continues to remain an important public health problem despite intensive efforts by the World Health Organization (WHO) to eradicate the disease by the year 2000 (WHO, 1988). Although the number of cases reported worldwide since 1988 has been reduced by 82%, the WHO estimates that up to 80000 cases of poliomyelitis occur each year, primarily in Asia and Africa (WHO, 1996). In 1993, the African region reported 15% of the global total of polio cases, with the majority of cases being reported from western and central Africa (WHO, 1994a). In southern Africa, several countries have reported relatively high levels of immunization coverage with three doses of oral polio vaccine in children under 1 year of age and, except for Namibia, very low or zero incidence of poliomyelitis. In 1994, the WHO suggested that southern Africa may be emerging as a polio-free zone (WHO, 1994b).
in the form of identification of wild-type virus infections are thus critical parts of the WHO strategy for eradication.

Polioviruses mutate rapidly upon passage in the human gastrointestinal tract, at a rate of approximately 1–2 base substitutions across the 7·5 kb RNA genome per week, or 1–2% per year (Nottay et al., 1981). The accumulation of changes in the viral genome over time has given rise to different genetic strains of wild-type polioviruses; these genetic variants have been termed genotypes and have been found to be distributed geographically (Rico-Hesse et al., 1987). A reasonable picture of the natural transmission and distribution of all three serotypes of wild-type polioviruses can be obtained by sampling as little as 150 bases, or 2%, of the viral genome (Rico-Hesse et al., 1987; Kew et al., 1990). Analysis of the extent of sequence divergence between poliovirus isolates associated with cases and outbreaks can provide information on the local, regional and global transmission pathways and patterns of circulation of wild-type polioviruses (Kew et al., 1990), and can lead to the detection of epidemiological links that cannot accurately be determined by any other means.

In this study we provide an overview of the past and present distribution of poliovirus type 1 genotypes in sub-Saharan Africa. Genetic relationships between viruses were determined by PCR amplification and sequence analysis of a 150 bp region across the VP1/2A junction (Rico-Hesse et al., 1987).

Methods

Viruses. Poliovirus isolates characterized in this study are listed in Table 1. All strains were isolated from clinical specimens of patients presenting with AFP or with a clinical diagnosis of poliomyelitis, or from contacts of such cases. Specimens from South Africa included those obtained during the outbreaks in 1982 in Gazankulu (Johnson et al., 1984) and in 1987–88 in KwaZulu-Natal (Schoub et al., 1992). The majority of isolates from Namibia were obtained from poliomyelitis cases during the 1993 epidemic (van Niekerk et al., 1994). Polioviruses from Central African Republic (CAR), Tanzania, Zambia and Zaire were isolated in WHO regional and national laboratories and were submitted to the National Institute for Virology (NIV) for intratypic characterization and/or sequence analysis. The Sabin 1 poliovaccine strain (P1/LSC2ab) was obtained from the Vaccine Unit, NIV, and was used as the reference strain in all alignments. Viruses were propagated on Hep2-C cells and typed using the microneutralization technique with antiserum pools supplied by the National Institute for Public Health and Environmental Protection, Bilthoven, The Netherlands. Intratypic differentiation between vaccine-like and wild-type strains was performed using the PCR assay described by Yang et al. (1991).

Reverse transcription, PCR and sequencing. In order to generate templates for sequencing, a 293 bp fragment, encompassing the 150 bp region used for genotype analysis, was amplified using poliovirus-specific primers as described previously (Chezzi, 1996). Briefly, RNA was extracted from infected tissue culture fluids by standard phenol–chloroform extraction procedures and amplified in a single-step reverse transcription PCR assay containing both the forward primer PVPCR2 (5′ GTC AAT GAT CAC AAC CCC AC) and reverse primer 2A (5′ AGG TCT CTA TTC CAC AT). Amplified products were purified from agarose gels using the Mermaid kit (BIO 101) and sequenced using the Sequenase PCR-product sequencing kit (US Biochemical) with [35S]-dATP. Templates were sequenced using primer 2A, and confirmed by repeat sequencing with primer PVPCR2. Sequencing products were resolved on 8% acrylamide gels containing 7 M urea, and visualized by autoradiography.

Analysis of sequence data. Sequence data was analysed using DNASTAR software (Hitachi). A matrix of genetic distances between strains was generated by performing pairwise comparisons of all sequences. Genetic relationships between poliovirus isolates were then determined by constructing a dendrogram, based on the calculated distances between strains, using the KITSCF programme from the PHYLIP phylogenetic inference package (Felsenstein, 1993). In the dendrogram, the distance along the abcissa to the node connecting any two strains is a measure of the extent of sequence divergence between those strains. A genotype was defined as a group of viruses showing at least 85% nucleotide sequence similarity across the 150 bp VP1/2A region (Rico-Hesse et al., 1987), and direct epidemiological links were inferred by a minimum nucleotide identity of 98% between strains (Rico-Hesse et al., 1987). Previously published sequences of the following isolates (Rico-Hesse et al., 1987), and direct epidemiological links were inferred by a minimum nucleotide identity of 98% between strains (Rico-Hesse et al., 1987). Previously published sequences of the following isolates (Rico-Hesse et al., 1987) were also used in the comparisons: 24115/KEN93, 24094/SUD93, 24122/EGY93, 08659/IND92, 21197/PAK92, 11236/EGY91 (Mulders et al., 1995), 1197/JOR78, 1177/KUW77, 8423/ISR88 (Rico-Hesse et al., 1987). Sequences for the following isolates were kindly made available to us by O. M. Kew, Centers for Disease Control, Atlanta, Ga., USA: 6747/SEN86, 5552/NIE93, 9475/ZAI, 6224/ZIM85, 5136/ETH93.

Results

Relationships between wild-type 1 polioviruses based on nucleotide sequence comparisons

More than 250 wild-type poliovirus strains isolated from countries in Africa from 1980–1996 were characterized by partial genomic sequencing of 150 bases across the VP1/2A junction region. For conciseness, results for isolates that were identical or very closely related in sequence are not presented. All analysed strains were found to be type 1. All mutations

---

**Table 1. Type 1 poliovirus isolates characterized by partial genomic sequencing**

<table>
<thead>
<tr>
<th>Country of origin</th>
<th>Year of isolation</th>
<th>Number of isolates</th>
<th>Supplied by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angola</td>
<td>1994</td>
<td>4</td>
<td>J. M. Morvan</td>
</tr>
<tr>
<td>CAR</td>
<td>1993</td>
<td>7</td>
<td>J. M. Morvan</td>
</tr>
<tr>
<td>Namibia</td>
<td>1982</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1993</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1994</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td>1980–1989</td>
<td>177</td>
<td></td>
</tr>
<tr>
<td>Tanzania</td>
<td>1995</td>
<td>3</td>
<td>E. M. Mpabalwani and H. Saijo</td>
</tr>
<tr>
<td>Zaire</td>
<td>1995</td>
<td>7</td>
<td>E. M. Mpabalwani and H. Saijo</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16</td>
<td>O. Tomori</td>
</tr>
</tbody>
</table>
observed were base substitutions, with a transition/transversion ratio of 2 (Fig. 1). However, within the major virus groups (genotypes), nearly all the substitutions between pairs of viruses were transitions (Fig. 1). A graphical representation of the sequence relationships between wild-type polioviruses in Africa, determined from the extent of genetic distance between strains, is shown in Fig. 2. It is important to note that the method used for constructing the dendrogram (KITSCH) imposes constraints on branch lengths in the dendrogram, so that the total branch length from the root of the tree to any strain is the same. As a result, the position of each node represents the average divergence between sequences connected by that node, and is not an absolute measure of the percentage divergence between any two strains.

Sequences group into seven major clusters, termed A to G, each representing a different genotype (Fig. 2). Genotype A was geographically the most widespread, comprising isolates from countries in western, south-western, central and southern Africa. Several smaller clusters, differing from each other by between 6% and 15%, were found within this genotype. Isolates obtained from central Angola in 1994 (5384/ANG94 and 1276/ANG94) and from the poliomyelitis outbreak in Zaire in 1995 (757/ZAI95 and 4097/ZAI95) exhibit 96% nucleotide identity between them and cluster together. All Zaire epidemic isolates exhibited between 98% and 100% nucleotide similarity. Strain 67/ZAM95, isolated in December 1995 from a case in Ikelenge/Mwinilunga in north-western Zambia, close to the border with Zaire, also fell within this cluster, and exhibited 99% genetic similarity to the Zaire outbreak isolates.

Isolates from the 1993 outbreak in Namibia form a separate cluster within genotype A. All isolates from the outbreak were closely related, displaying a maximum sequence divergence of 2%. The closest relatives were from Angola—isolates from the Namibian epidemic (1405/NAM93 and 1177/NAM93) and from central Angola (5384/ANG94 and 1276/ANG94).
displayed approximately 94% sequence identity. A strain from 1992 from the west African country of Togo shared 92% sequence identity with Namibian outbreak isolates. Strains isolated from northern Namibia early in 1995 (72/NAM95 and 818/NAM95) belong to a third cluster within genotype A. These strains diverged by 12% from strains from the 1993 outbreak in Namibia, but displayed a very high percentage sequence similarity (99%) with isolates made from cases in southern Angola late in 1994 (1497/ANG94).

Strains isolated in 1994 from CAR (76/CAR94 and 50/CAR94) make up a fourth cluster within the A genotype. Eleven isolates from a total of 16 analysed belong to this cluster. Sequence variation within this cluster was limited to a maximum of 2%. The highest degree of identity with other strains within genotype A, although limited to 88%, was observed with strains from central Angola from 1994.

A separate cluster within genotype A consisted of isolates from the outbreak in Senegal in 1986 (6747/SEN86; WHO, 1987), isolate 5552/NIE93 from Nigeria, a 1982 isolate from Namibia (4926/NAM82) and five isolates obtained in South Africa between 1982 and 1985.

Genotype B consists of strains from CAR from 1993 (1/CAR93 and 17/CAR93). This genotype includes strains from the same period from Kenya (24115/KEN93), Sudan (24094/SUD93), Egypt (24122/EGY93) and Ethiopia (5136/ETH93), and older strains from India (08659/IND92) and Pakistan (21197/PAK92).

Strains belonging to a second independent genotype (genotype G) were co-circulating simultaneously with those belonging to genotype B in CAR in 1993. The maximum sequence divergence between isolates 20/CAR93 and 16/CAR93 was approximately 3%. Included in this group are older isolates from Egypt (11236/EGY91) and the Middle East (8423/ISR88).

Recent poliovirus strains isolated in 1995 from Tanzania (564/TAN95 and 1199/TAN95) and Zambia (50/ZAM95) belong to genotype C, which appears to be geographically restricted to eastern and southern Africa (Fig. 3). Isolates from Tanzania were made from geographically dispersed cases, and all isolates were very closely related, with a maximum sequence divergence of 1-5%. Six strains isolated from cases in Lusaka, Zambia later in 1995 exhibited a maximum of 2% sequence divergence from Tanzanian strains. A strain from Zaire (9475/ZAI) also falls within genotype C, exhibiting a maximum sequence similarity of 87% with isolates from
Poliovirus genotypes in Africa

Fig. 4. Predicted amino acid sequences of the 150 bp VP1/2A interval used for determination of genetic relationships between poliovirus strains. Dots represent amino acids which are identical to those of Sabin 1. Country abbreviations are the same as those defined in the legend to Fig. 1.

Tanzania and Zambia. The exact date of isolation of this strain, except for the fact that it was definitely before 1987, is unknown, and no additional early isolates from Zaire were available for analysis.

Genotypes D, E and F comprise isolates from South Africa. These three genotypes cover the same geographical regions, but are separated temporally. Strains belonging to genotype D appeared in South Africa early in 1985 and remained in circulation until 1989. All poliovirus strains isolated in the country since 1986 belong to this genotype. The earliest strain belonging to genotype D (427SOA85) was isolated from a case in Nelspruit, in the north-eastern area of the country, and the closest relative (4% divergence) was a 1985 strain from Zimbabwe (6224ZIM85). Strains isolated throughout South Africa during the 3 year period between 1985 and the end of 1987 displayed approximately 5–6% sequence divergence. The poliomyelitis epidemic in 1987–88 in Kwazulu-Natal was associated with poliovirus strains belonging to genotype D and, as expected, the majority of epidemic isolates (359/ SOA88, 221/ SOA88 and 364/ SOA88) were very closely related (maximum sequence divergence 2%). Isolates obtained during 1989 were also very closely related to epidemic strains (divergence 2%).

Genotype E consists of strains isolated in South Africa from 1980 until 1985, and includes isolates from the Gazankulu outbreak of 1982. Sequence heterogeneity within this genotype was relatively high, with divergence between strains reaching 11%. Outbreak isolates (1532/ SOA82 and 2388/ SOA82) were very closely related (maximum divergence 2%) and clustered on their own within the genotype. No strains related to those from South Africa were identified in other African countries, thus making genotype E unique to South Africa. This genotype was present in South Africa since at least 1980 (1072/ SOA80 and 1897/ SOA80), and continued to be transmitted until late 1985.

Genotype F comprises strains which circulated in South Africa between 1983 and 1985. The minimum divergence between these isolates was 3% for isolates obtained during the same year (2/ SOA83 and 140/ SOA83). Later isolates from 1984 (2972/ SOA84) and 1985 (442/ SOA85) diverged by as much as 6% from the 1983 isolates. The closest relatives to these strains (85–88%) were older strains isolated in 1977–78 from cases in the Middle Eastern countries of Jordan (1197/ JOR78) and Kuwait (1177/ KUW77). Isolates within genotype F are completely unrelated (> 20% divergence) to the more recent Middle Eastern strains within genotype G.

Amino acid substitutions in the VP1/2A region

The predicted amino acid sequences encoded by the 150 bp VP1/2A sequences are shown in Fig. 4. The majority of the nucleotide differences between strains were silent, producing synonymous codons. The South African strains belonging to genotype E, with the exception of isolates from 1984 and 1985, had an amino acid composition identical to that of Sabin 1. The newer strains within this genotype (710/ SOA84 and 855/ SOA85) contained an arginine to alanine substitution at position 8 of protease 2A.

The most common substitution occurred at residue 20 of VP1; all strains within genotypes A, C, D, F and G substituted an alanine for a threonine at this position. Strains within genotype G contained an additional substitution of a threonine for serine at position 295 of VP1. One strain within this genotype, 20/ CAR93, contained additional substitutions of an arginine for alanine at residue 276 of VP1 and a serine to tyrosine substitution at residue 19 of protease 2A. With the exception of 17/ CAR93, which also exhibited the alanine for threonine-292 substitution, all strains within genotype B had a single substitution of a tyrosine or histidine for phenylalanine at residue 2 of protease 2A. A leucine for phenylalanine substitution at this residue was seen for isolates within genotype D. Strains 144/ SOA82 and 5552/ NIE93 in genotype A also contained the leucine for phenylalanine-2 substitution, whereas strain 1276/ ANG94, also in genotype A,
contained a cysteine for phenylalanine substitution at this position.

Several strains had other substitutions scattered within the 50 aa region under investigation. Two isolates, 9475/ZAI and 20/CAR93 differed from Sabin 1 by as many as four residues (8%).

Discussion

Partial genomic sequence analysis was employed to study the molecular epidemiology and patterns of circulation of wild-type 1 polioviruses in Africa. The 150 bp region spanning the VP1/2A junction was chosen for sequence analysis, as it has been shown that although this region represents only 2% of the poliovirus genome, this is sufficient to obtain a reasonable picture of the natural distribution and transmission of wild-type polioviruses (Rico-Hesse et al., 1987). This region codes for part of the viral capsid protein VP1 and for part of 2A, which has a known protease function, so that differences in mutation rates for structural and non-structural proteins, if any, can be determined. Analysis of the VP1/2A region has been used to study the molecular epidemiology of wild-type polioviruses of all three serotypes from the European, Middle Eastern and Indian subcontinent regions (Mulders et al., 1995), Pakistan (Huovilainen et al., 1995), the former Soviet Union (Lipskaya et al., 1995) and China (Zheng et al., 1993), and a large sequence database containing representative strains of the poliovirus genotypes circulating worldwide has been developed.

With the exception of wild-type 2 isolated in South Africa in 1980 (data not shown) and wild-type 3 from CAR (Gouandjika et al., 1995), the majority of wild-type polioviruses circulating in Africa during the past 15 years have been found to be type 1. No evidence of recombinant genomes, characterized by VP1 and 2A sequences derived from two different genotypes, was found. The rates of mutation for sequence intervals coding for VP1 and 2A were found to be uniform.

To date, seven major polio type 1 genotypes have been found in sub-Saharan Africa (Figs 2 and 3). Genotype A, which has been present since at least 1980, has the widest distribution, covering western, south-western, central and southern Africa. This genotype may well have been circulating still earlier in Africa, since older (1970–77) isolates from Senegal and Cameroon (Rico-Hesse et al., 1987) also just fall within this genotype. Within genotype A are several smaller clusters which show a high degree of diversity from each other and may represent emerging new genotypes. They appear to be segregated geographically, indicating independent sustained circulation of these lineages in different countries. One cluster includes isolates from central Angola and Zaire – the high degree of sequence identity between isolates from the two countries (96%), and the geographical proximity of the countries suggests possible importation of the epidemic strains into Zaire from Angola 2–3 years earlier. The 99% sequence similarity between strains from Zaire and a strain isolated from Zambia several months later is indicative of a definitive epidemiological link between cases, and evidence for direct transmission of this genotype into Zambia. The absence of any further isolates belonging to this genotype from Tanzania may indicate that further transmission of this strain in the country was successfully interrupted by effective control measures. A common link between cases from central Angola and isolates from the 1993 epidemic in Namibia is also evident, possibly indicative of importation of the epidemic strain from Angola into previously polio-free Namibia, although the 6% divergence between strains isolated only months apart points to the introduction and silent transmission of this strain in Namibia 3–4 years earlier. The earlier presence of genotype A within Namibia is evidenced by the isolation of 4926/NAM82, which diverged from the later epidemic isolates by 14%. Since the availability of sequence data for strains isolated in Namibia prior to 1993 is limited to that for 4926/NAM82, it is not possible to determine whether this strain represents a progenitor of the later 1993 epidemic isolates, or a strain imported into Namibia from western Africa.

The presence, within a single country, of several reservoirs capable of independent transmission of polioviruses for prolonged periods of time, as evidenced by the concurrent circulation of more than one poliovirus genotype, was seen in CAR and South Africa. In CAR, two poliovirus genotypes, B and G, were co-circulating during 1992–93. Viruses belonging to genotype B represent strains most likely imported into CAR from Sudan. The origins of genotype B, which is equivalent to genotype 4 described by Mulders et al. (1995), are in the Indian subcontinent; strains belonging to this genotype have been associated with several outbreaks and have circulated for over 10 years in Europe, the Middle East and the Indian subcontinent (Mulders et al., 1995; Huovilainen et al., 1995). Genotype G represents strains imported from the Middle East, probably via Egypt. This genotype has also been previously described, associated with cases in Middle Eastern countries (Mulders et al., 1995). The high degree of diversity (9%) between isolates from CAR and Egypt, separated by a time interval of only 2 years, suggests continued silent transmission of this genotype over a period of several years.

In South Africa, three independent poliovirus genotypes, A, F and E were in circulation between 1982 and 1985. The total of only five strains belonging to genotype A, isolated from a wide geographical area in the country and displaying 87% sequence similarity to strains isolated in west Africa, suggests that these strains were not originally endemic to the country, but represent imported strains, most probably from countries in western Africa. Similarly, strains belonging to genotype F represent importation from the Middle East. The sequence heterogeneity between the South African strains belonging to these genotypes is suggestive of progenitor infections in the late 1970s. The absence of any epidemiological relationships between strains from genotypes F and G,
circulating in the Middle East in the 1970s and late 1980s respectively, suggests that genotype G was not derived simply by evolution of genotype F, but represents an independent genotype. Genotype E was endemic in South Africa since at least 1980. The high degree of sequence divergence between South African strains within each genotype suggests widespread circulation and sustained endemic transmission of individual strains along separate pathways for extended periods.

Simultaneous co-circulation of more than one genotype in regions with sub-optimal health services has also been reported from China (Zheng et al., 1993), the former Soviet Union (Lipskaya et al., 1995), the Middle East and Pakistan (Mulders et al., 1995).

Displacement of endemically circulating genotypes by an imported genotype, favoured by conditions of low vaccine coverage allowing transmission of the imported strain amongst pockets of susceptible individuals (Rico-Hesse et al., 1987), occurred both in CAR and South Africa. This condition has also been reported in developing countries in South America (Rico-Hesse et al., 1987). In CAR, genotype A displaced the imported B and G genotypes in 1993, although the high degree of sequence divergence between strains of this genotype from CAR and isolates from other African countries excludes any direct epidemiological links. Genotype D was introduced into South Africa in 1985, most probably from countries north of the border, and displaced the endemically circulating genotypes, as evidenced by the subsequent isolation of genotype D strains only. In contrast to the high degree of divergence (up to 11%) seen between strains belonging to genotype E, strains belonging to genotype D displayed less sequence heterogeneity (maximum 6%), and no apparent clustering other than that displayed by epidemic isolates.

The identification of a definitive epidemiological link between cases from Tanzania and Zambia illustrates the resolving power of molecular epidemiology in establishing otherwise unrecognized links. The high degree of sequence similarity between isolates from these countries is indicative of direct importation of the Tanzanian strain into Lusaka. The isolation in Lusaka of strains displaying highly similar nucleotide sequences (99–100%) suggests rapid epidemic transmission amongst inadequately immunized individuals. The origin and past distribution of genotype C, to which the Tanzanian and Zambian isolates belong, cannot be accurately determined because of the lack of earlier specimens from both those and neighbouring countries. It appears to be confined to south-eastern Africa, although the 87% similarity with isolate 9475/ZAI, isolated at least 10 years earlier from Zaire, suggests that this might be the progenitor strain and that earlier coverage by this genotype may have included central Africa.

Strains from genotype C and D exhibit an average of 84% sequence similarity (sequence divergence of 14%, however, is seen between 50/ZAM95 and 6226/ZIM85, and between 50/ZAM95 and 1170/SOA88). Although this exceeds the 15% divergence limit set for identification of relationships based on epidemiological grounds, it is possible that the two genotypes, which are separated temporally by about 10 years and which cover neighbouring geographical areas, may represent separate evolution of the same progenitor genotype along two different transmission pathways. Results of bootstrap analysis of the sequence data in construction of the dendrogram (not shown) indicate that the branching pattern connecting genotypes C and D is significant, occurring in >95% of trees sampled.

Accurate determination of relationships between type 1 polioviruses based on their amino acid sequences was not possible, as most of the base differences between strains were silent and produced synonymous codons. Consistent with published data (Rico-Hesse et al., 1987), the most frequent substitution was alanine for threonine at position 292 of VP1 (residue 20; Fig. 4); this mutation appears to be common for all strains, except those originating from the Indian subcontinent and those belonging to genotype E, endemic to South Africa between 1980 and 1985. Close analysis of the amino acid sequences revealed that there appears to be a correlation between the amino acid substitutions and specific genotypes. The double substitution of alanine for threonine and threonine for serine at residues 20 and 23 of VP1 respectively was found only in strains that fell into genotype G, endemic to the Middle East. Analysis of the published amino acid sequences of wild-type 1 strains from a wider geographical region (Rico-Hesse et al., 1987) revealed that strains obtained between 1980 and 1984 from Venezuela, introduced into the country from the Middle East, also contained the same two substitutions. Isolates obtained 10 years apart, from CAR and Venezuela, exhibit 86–88% sequence similarity (results not shown), indicative of their relatedness to the same progenitor strain from the Middle East. Consistent with published data for type 1 strains from Pakistan (Huovilainen et al., 1995) and from Tajikistan, north of the Pakistan border (Lipskaya et al., 1995), substitution of a tyrosine or histidine for phenylalanine at position 2 of 2A was seen for strains belonging to genotype B, related to isolates from the Indian subcontinent. A unique leucine for phenylalanine substitution at the same position was evident for strains belonging to genotype D, circulating in southern Africa until 1989.

The limited number of amino acid substitutions in the VP1/2A region may indicate a requirement for polypeptide structure conservation in both the carboxyl terminus of VP1 and the amino terminus of protease 2A. There does, however, appear to be a tendency for the conservation of unique substitutions associated with certain genotypes, although the significance of this observation is unclear. None of the mutations observed were found to occur at sites shown, by mutational analysis of the substrate recognition determinants of protease 2A, to be important for autoproteolytic cleavage at
the VP1/2A junction (Hellen et al., 1992); variation at residue 2 of 2A, seen for isolates in genotypes B and D, has been found to be tolerated, and the critical threonine and glycine residues at positions 29 of VP1 and 1 of 2A were conserved for all strains examined.

In this study we have described the general distribution and molecular epidemiology of wild-type 1 poliovirus genotypes in sub-Saharan Africa. Although by no means complete, due to the unavailability of isolates from many countries, the results provide a comprehensive overview of the past and present wild-type 1 poliovirus circulation across most of the continent. Of the seven genotypes described here, three, previously circulating in southern Africa (genotypes D, E, and F) appear to have been eliminated, and two imported genotypes (B and G) appear to have been displaced, at least certainly from CAR. The distribution of genotype C appears to be limited to south-eastern Africa, and to date it has only been associated with a relatively small number of cases. Genotype A continues to circulate endemically throughout most of west and central Africa.

We would like to thank J. M. Morvan (Central African Republic), E. M. Mpabalwani (Zambia), M. Saijo (Zambia) and O. Tomori (WHO) for providing isolates included in this study, and O. M. Kew (CDC) for providing some unpublished sequence data. Reisolation of many of the poliovirus strains was performed by E. Maselesele (JIV). This work was supported in part by grants from the Poliomyelitis Research Foundation and the South African Medical Research Council.

References
Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle, USA.

Received 12 August 1996; Accepted 19 December 1996