Twenty Years of Prospective Molecular Epidemiology in Senegal: Changes in HIV Diversity

DONALD J. HAMEL,1 JEAN-LOUIS SANKALÉ,1 GEOFFREY EISEN,1 SEEMA THAKORE MELONI,1 CHRISTOPHER MULLINS,1 AISSATOU GUEYE-NDIAYE,2 SOULEYMANE MBoup,2 and PHYLLIS J. KANKI1

ABSTRACT

Over a 20-year period we have observed the dynamics of HIV-1 subtypes and HIV-2 infection in a prospective cohort of registered female sex workers (FSW) in Dakar, Senegal. Prevalence and incidence rates for HIV-1 and HIV-2 are described from 290 seroprevalent and 193 seroincident subjects who were among the 3910 women enrolled between 1985 and 2004. We report a significant decrease of HIV-2 prevalence in the cohort, parallel to the introduction and rise of HIV-1 infection. In 328 HIV-1-infected women, a 385-bp C2–V3 fragment of the envelope gene was sequenced and classified into the following subtypes or recombinant forms: 239 (72%) were subtype A [of which 180 (55%) were CRF02_AG and 53 (16%) were A3], 10 (3%) were B, 12 (4%) were C, 11 (4%) were D, 18 (6%) were G, 24 (7%) were CRF06_cpx, and 7 (2%) were CRF09_cpx. We found an increasing proportion of CRF02_AG over many years, but recently subsubtype A3 has overtaken CRF02_AG, with the largest proportion of new infections. The predominance of existing HIV-1 subtypes did not preclude the emergence and increase of other closely related subtypes or recombinant forms. This 20-year prospective serological and sequence analysis of HIV viruses reveals a complex and changing HIV epidemic in Senegal.

INTRODUCTION

HISTORICALLY, THE FIRST HIV-1 VIRUSES isolated and characterized were from the United States and Europe. In the early 1990s, it was discovered that multiple HIV-1 subtypes existed, but surprisingly the U.S. and European epidemics resulted from a single HIV-1 subtype, subtype B. It is now recognized that HIV-1 can be classified into three groups (M, N, O) and that the M group consists of at least nine different subtypes (A, B, C, D, F, G, H, J, and K) and circulating recombinant forms (CRFs). Most of the diverse HIV-1 subtypes and HIV-2 have been described in sub-Saharan Africa where they are associated with the largest proportion of HIV infections worldwide. It is still poorly understood why these distinct subtypes and their epidemics have emerged or how they will evolve over time. Differences in transmission or disease potential between HIV-1 subtypes have been described, but their association with the emergence or contribution to the dynamics of the HIV epidemics has not been well studied.1–3

Since 1985 we have conducted a prospective study of registered female sex workers (FSW) in Dakar, Senegal. HIV-2 was first described in this cohort of FSWs and the initial prevalence rate for HIV-2 in infection in 1985 was 8%.4 By contrast, HIV-1 was first detected in Senegal in 1985, and since then has remained at a constant prevalence rate of about 1% in the general population,5,6 although HIV-1 infection in FSWs had increased to over 13% by 2003.

Globally, most areas have just one or two HIV-1 subtypes geographically. We, along with others, have shown that the predominant HIV-1 subtypes circulating in West Africa are subtype A and AG recombinants, including the circulating recombinant form CRF02_AG.7–13 New subtypes, subsubtypes, and CRFs continue to be identified, and the issue of recombination has become an important consideration in tracking the global spread of HIV. Most studies that have analyzed subtype distribution are based on a cross-sectional design. Over a 20-year period, we enrolled nearly 4000 women and sequenced all available HIV-1 samples to study the dynamics of HIV-1 group
M subtypes. Our prospective design has allowed us to document the introduction and dominance of specific subtypes and CRFs through the analysis of incident infections.

MATERIALS AND METHODS

The study focuses on an open cohort of female commercial sex workers who were able to enroll or leave the cohort throughout the study period. Samples were obtained during regular biannual visits to the Institut d’Hygiène Sociale clinic in Dakar, Senegal, which are a requirement for legal registration of the sex workers. At these visits, women were provided with a clinical examination as well as treatment for any sexually transmitted diseases. Women who gave informed consent were enrolled into our study group. Serology was performed using immunoblot, as well as synthetic peptide assays, to diagnose infection with HIV-1, HIV-2, or HIV dual infection (HIV-D). The detailed protocols for study recruitment and the cohort design have been described previously. Seroincident HIV-1 cases were defined as those who were HIV negative at enrollment and subsequently seroconverted to HIV-1. Time of infection, or seroconversion date, was calculated as the midpoint between the last seronegative sample date and the first seropositive sample date. HIV-1-infected individuals who were also HIV-2 infected, or HIV dually infected, were included in this study.

Blood samples were collected from the women upon enrollment and at scheduled visits, and separated into plasma and peripheral blood mononuclear cell (PBMC) fractions. For HIV-1-positive subjects, proviral DNA was extracted from PBMCs (Qiagen Blood DNA Kit; Qiagen Inc., Chatsworth, CA). We performed a nested polymerase chain reaction (PCR) of a 385-bp fragment of the envelope gene (C2–V3) using previously described primers and conditions. The PCR product was then sequenced using the second round PCR primers, KK30 and KK40. Products were cloned when necessary, using the pCR2.1 vector (T/A Cloning; Invitrogen, San Diego, CA). Sequence reactions were performed by either dye terminator cycle sequencing using Taq polymerase, or by the Big Dye Terminator v1.1 system (both Applied Biosystems Inc., Foster City, CA). For incident cases, DNA samples were selected as the first available after seroconversion, typically from the same bleed date as the first seropositive result.

Alignments of sequences were performed using the Clustal multiple alignment software package (Clustal W 1.6). Minor manual adjustments were made where necessary to accommodate reading frames. Phylogenetic analysis was performed using the neighbor-joining method, and reliability was estimated from 1000 bootstrap resamplings. Representative sequences from the HIV database were included in the analysis.

The nucleotide sequences were submitted to GenBank (accession numbers AF085284–085327, AF020819–020827, AF526650–526876, and DQ323178–323400).

RESULTS

Annual HIV prevalence and incidence

From 1985 to 2004, we enrolled 3910 women into the Dakar FSW cohort; of these, 290 were HIV-1 seropositive (sero-prevalent) on entry and 193 seroconverted to HIV-1 (seroincident) during observation in the cohort. Based on the serology screening results, we calculated the annual prevalence and incidence rates (IR) for HIV-1, HIV-2, and HIV-D. HIV-1 prevalence has climbed steadily since its introduction in 1985 to just below 10% in 1993 (Fig. 1A). HIV-1 prevalence moderated from 1993 to 1998, and then continued to rise again to 13.8% in 2003. HIV-2 decreased from the range of 8–11% between 1985 and 1995, to 5.5% in 2003. HIV-D increased from 0.3% in 1985 to 3.2% in 1995, and then decreased to 1.8% in 2003.

The annual incidence rate for HIV-1 increased from 0 in 1985 to 2.5/100 person years of observation (PYO) in 1992, while HIV-2 and HIV-D have decreased to less than 0.3/100 PYO each since 1999 (Fig. 1B). The incidence rate for HIV-1 has also remained higher than that of HIV-2 and HIV-D since 1990, in the range of 0.9–2.8/100PYO.

Subtype summary and phylogenetic analysis

We obtained subtype data for 151/193 (78%) seroincident individuals identified, and similarly subtyped 177/290 (61%) sero-prevalent individuals. The subtyped seroincident and sero-prevalent groups together comprise a total of 328 subjects.

FIG. 1. (A) Annual prevalence for HIV. (B) Annual incidence rates for HIV.
(Table 1). Of the 328 subjects for whom we present subtype data in this paper, 223 represent novel sequences that have not been published or analyzed previously.

Assignment of specific subtypes was determined by phylogenetic analysis of all sample sequences with reference subtype sequences using the neighbor-joining method. Due to the large number of samples analyzed here, we have presented representative data in two phylogenetic trees, one for the prevalent group (Fig. 2A) and one for the incident group (Fig. 2B). All sequences, with the exception of three unclassifiable subtype A viruses, clustered definitively with reference subtypes with bootstrap values between 53 and 100%. With the small size of the fragment sequenced, it can be difficult to achieve high resolution between closely related subtypes such as subtype G and CRF06_cpx or between CRF02_AG and sub-subtype A3. Subsubtype A3 has only recently been identified and is not yet included in subtype reference alignments published by the Los Alamos National Laboratories.

A majority of the women, 72% (239/328), were infected with subtype A viruses, and of those most 55% (180/328) were CRF02_AG. CRF06_cpx contributed 7% (24/328), subtype G 6% (18/328), and all other subtypes and recombinant forms contributed less than 5% (15/328) each to the overall total in the cohort.

In the 151 individuals with documented new infection (seroincident), the contributions from each subtype were determined for each calendar year. From 1986 to 1990 incident cases included subtypes D, C, A1, and CRF02_AG. Subsubtype A3 was introduced into the cohort in 1991, CRF06_cpx in 1992, subtype G in 1993, and CRF09_cpx in 1994. In 1999, all incident cases were subtype A, with 71% (12/17) from CRF02_AG and 24% (4/17) from subtype A3.

Likewise, the distribution of each subtype was determined for every year among the seroprevalent cases. We documented the first prevalent case of subsubtype A3 in 1988, 3 years before the first incident case of subsubtype A3 was observed. The first detection of CRF06_cpx was in 1990, 2 years before the first incident case was observed. Also, the first evidence of subtype G among the prevalent was in 1992, 1 year before the first incident case.

**Subtype trends**

A timeline was created for the 151 incident cases for which we have subtype data (Fig. 3). The timeline graphically illustrates the introduction of subsubtype A3, CRF06_cpx, and subtype G into the incident cohort. Upon the introduction of subsubtype A3 in 1991, it has increased to represent 23% of new incident infections over the 20-year study period. Within this incident group we documented the diagnosis of AIDS in 13 women, also graphically illustrated on the timeline.

To better resolve trends of specific subtypes over time, and to minimize the impact of small annual fluctuations, we created tables collapsing data into 5-year intervals for both the HIV-1 seroincident cases (Fig. 4A) and for the HIV-1 seroprevalent cases (Fig. 4B).

Among the HIV-1 seroincident cases, there was no evidence of CRF06_cpx during the period from 1985 to 1989, but this CRF represented 6% of infections from 2000 to 2004 (Fig. 4A). We also documented the increase in new HIV-1 infections by subsubtype A3 from 0% in 1985–1989 to 41% in 2000–2004.

For the seroprevalent group, subtype G was not observed during the 1985–1991 period, but this subtype contributed 22% of HIV-1 cases entering the cohort from 2000 to 2004 (Fig. 4B). The dominant subtype remained CRF02_AG throughout the study period, representing 50% from 1985 to 1989 and 44% from 2000 to 2004.

**Epidemiologic analyses**

For the total 483 women who tested HIV positive during the study period, we were unable to subtype 155 (32%) due to lack of an available DNA sample, predominantly from the study years of 1985–1986 and 2002–2004. The low numbers of subtype data from recent years should not be inferred to necessarily represent a decline in incidence or prevalence. Analyses were performed to examine sampling bias as a result of the open

### Table 1. Summary of HIV-1 Subtype Results

<table>
<thead>
<tr>
<th>Subtype</th>
<th>Seroincident cases</th>
<th>Seroprevalent cases</th>
<th>Total cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>115</td>
<td>124</td>
<td>239 (72)</td>
</tr>
<tr>
<td>A1</td>
<td>2</td>
<td>1</td>
<td>3 (1)</td>
</tr>
<tr>
<td>A3</td>
<td>32</td>
<td>21</td>
<td>53 (16)</td>
</tr>
<tr>
<td>A unclassified recombinant</td>
<td>1</td>
<td>2</td>
<td>3 (1)</td>
</tr>
<tr>
<td>CRF02_AG</td>
<td>80</td>
<td>100</td>
<td>180 (55)</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>7</td>
<td>10 (3)</td>
</tr>
<tr>
<td>C</td>
<td>9</td>
<td>3</td>
<td>12 (4)</td>
</tr>
<tr>
<td>D</td>
<td>7</td>
<td>5</td>
<td>12 (4)</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>2</td>
<td>3 (1)</td>
</tr>
<tr>
<td>G</td>
<td>3</td>
<td>16</td>
<td>19 (6)</td>
</tr>
<tr>
<td>J</td>
<td>0</td>
<td>2</td>
<td>2 (1)</td>
</tr>
<tr>
<td>CRF 06</td>
<td>7</td>
<td>16</td>
<td>23 (7)</td>
</tr>
<tr>
<td>CRF 09</td>
<td>6</td>
<td>2</td>
<td>8 (2)</td>
</tr>
<tr>
<td>Total</td>
<td>151</td>
<td>177</td>
<td>328</td>
</tr>
</tbody>
</table>
FIG. 2. (A) Representative neighbor-joining radial phylogenetic tree of HIV-1 prevalent cases. Reference strains representing group M subtypes and CRFs, designated by ●, are as follows: A1_SE.94.SE7253, A1_UG.92UG037, B_FR.83.HXB2, B_US.90.WEAU160, CRF02AG_FR.91.DJ264, CRF02AG_GH.G829, C_BW.96BW0502, C_BR.92BR025, CRF09_TM7808, CRF06_CX.ML.95ML84, CRF06_CX.SN.97SE1078, D_CD.83.ELI, D_CD.84ZR085, F1_BE.93.VI850, F1.FR.96.M411, G_NG.92NG083, G_SE.93.SE6165. (B) Representative neighbor-joining radial phylogenetic tree of HIV-1 incident cases. Reference strains representing group M subtypes and CRFs, designated by ●, are as follows: A1_SE.94.SE7253, A1_UG.92UG037, B_FR.83.HXB2, B_US.90.WEAU160, CRF02AG_FR.91.DJ264, CRF02AG_GH.G829, C_BW.96BW0502, C_BR.92BR025, CRF09_TM7808, CRF06_CX.ML.95ML84, CRF06_CX.SN.97SE1078, D_CD.83.ELI, D_CD.84ZR085, F1_BE.93.VI850, F1.FR.96.M411, G_NG.92NG083, G_SE.93.SE6165.
FIG. 3. HIV-1 seroincident data timeline representing the distribution of infections among the FSWs in our cohort. The line before the infection date represents an individual's negative person-years of observation, and the color line after the infection date represents the positive person-years of observation. The start of each line indicates the first bleed date for each subject, and the end of line indicates the last bleed date.
FIG. 4. (A) Incidence rates of subtyped HIV-1, with 95% confidence intervals. Incidence rates for HIV-1 and individual subtypes are calculated annually before being compiled into 5-year groups. Each person contributes only once toward incidence. (B) Prevalence of subtyped HIV-1, with 95% confidence intervals. Prevalence rates for HIV-1 and each subtype are calculated annually before being compiled. Mean annual prevalence is the sum of the prevalence for each year divided by the number of years for each group.
Molecular Epidemiology of HIV in Senegal

cohort design, which could affect our overall conclusions. Pearson’s chi-square tests revealed that a higher proportion of individuals of Senegalese nationality ($p > 0.001$) and HIV-D ($p > 0.001$) were likely to be untyped, compared to the non-subtyped individuals. Similarly, previous analyses have shown that Ghanaian women were more likely to drop out of the cohort and less likely to have been subtyped. Therefore we must exercise some caution in extending our conclusions about subtype dynamics for groups with nationality other than Senegalese. Our analysis found no significant difference comparing the subtyped to the nonsubtyped with respect to religion, education level, marital status, or age at which the individual was diagnosed as HIV positive.

**DISCUSSION**

This study describes the molecular epidemiology of the HIV-1 subtypes in Senegal over the past 20 years. We have documented new observations of HIV-1 subtype dynamics in this region; our data are also consistent with cross-sectional prevalence studies. Our observations on subtype introduction and dynamics are further confirmed by the longevity and prospective design of the study. An overview of the annual incidence data for HIV (Fig. 1B) reveals a decreasing rate for HIV-2 and HIV-D infections, but a fairly stable rate of new HIV-1 infections. The incidence rate for HIV-1 has remained approximately 1.5% since 1990. Our analysis of the annual prevalence (Fig. 1A) demonstrates a decrease of HIV-2 and HIV-D infections. However, we have seen an increase in HIV-1 infection, after moderation between 1992 and 1999, reaching a prevalence of 13.8% in 2003 and continuing to increase in 2004.

This cohort consists of women, self-identified as registered sex workers, with higher rates of HIV infection than other populations in Senegal, such as blood donors or pregnant women visiting antenatal clinics. Previous analyses have demonstrated that the women enrolled in this cohort are representative of the registered sex worker population with regard to potential confounding risk factors including nationality, age, years of registered prostitution, and distributions of HIV serostatus. These studies of the cohort have also shown the predominant nationality of the women enrolled to be Senegalese (75.2%), or Ghanaian (16.7%), with Guinea-Bissau and other nationalities comprising the remainder (8.1%).

The initiation of this study early in the Senegal HIV-1 epidemic and the prospective study design carried out for a 20-year period are unique, and provide important new information on the epidemics of both HIV-1 and HIV-2, as well as the introduction of distinct HIV-1 subtypes in this region. We have described and characterized the entry and different rates of spread of subtype G, subtype A3, and CRF06_cpx. The rapid increase in subtype A3 infection suggests a “fitness” advantage that is worthy of further study. Sarr et al. have previously described a larger proportion of subtype A3 infections in HIV/HIV-2 dually infected individuals in this cohort, which led to the possibility that this variant might have a transmission advantage. Our subtype classification is somewhat limited by the subgenomic sequence analysis, which does not allow a definitive conclusion of subtype throughout the genome. For the 328 individuals with sequence data presented here, we have also performed gag gene sequencing on a subset of 44 (data not shown) that was classified as CRF02_AG or subtype A3 based on envelope region sequence. In phylogenetic analysis 42 of 44 (95%) of these gag sequences clustered with the same subtype as the env sequences. In the two discordant samples, a recombination event may have occurred, as one sample clustered with CRF02_AG in env and subtype A3 in gag, and the other sample clustered with subtype A3 in env and CRF02_AG in gag. Therefore the majority of samples tested revealed a similar sequence subtype in two regions, lending support for our subtype designations by env subgenomic sequencing. However, we note that the proportion of recombinant viruses may be underestimated based on our methodology.

HIV-2 was the dominant virus in 1985, when HIV-1 was first entering this population. Between 1985 and 2004, HIV-2 prevalence dropped from 8% to 4%, a 50% decrease over 20 years. Concurrently, HIV-1 prevalence increased from 0 to 13% during the same time period. Anderson and May discussed predictions for the interaction of HIV-2 and HIV-1 on a population level. Their prediction that HIV-2 would decrease in the face of an increasing HIV-1 prevalence appears to be supported by our data. The attenuated rate of HIV-1 infection in a high risk group of sex workers would also suggest that the interaction of the HIV viruses in this population may inhibit explosive increases in HIV-1 that have been described in similar HIV-1 studies of sex workers. In addition, the registration and health care system for these women supported by a strong government HIV/AIDS control, prevention, and treatment program may have contributed to the decline or stabilization of infection rates for both viruses.

Our prospective study design has reduced the risk of a sampling bias that would, for example, be of concern in a cross-sectional study focusing on hospitalized patients. Examining this cohort over time has allowed us to describe the rapid rise of subtype A3 since the late 1990s. Our timeline of incident HIV infection reveals the emergence of subtype A3 in 1991, CRF06_cpx in 1992, and subtype G in 1993. Similarly, analysis allowed the identification of recombinant CRF09_cpx, which entered the incident group in 1994 and was first described in this cohort. Additionally, these data support and confirm the predominance of recombinant subtype CRF02_AG in West Africa, as documented in both the HIV-1 incident and prevalent groups of our study.

The continent of Africa bears the world’s highest HIV infection burden (UNAIDS, 2004), as well as the highest diversity of HIV-1 subtypes. While HIV-1 subtype B continues to remain the most characterized strain, other subtypes have perhaps greater importance, for example, subtype C infection is responsible for over 50% of all infections worldwide. Previous studies have suggested higher transmission risk from subtype CRF01_AE than subtype B.
data significantly add to the knowledge of the molecular epidemiology of HIV in Senegal, and advances the goal of understanding the importance of multiple subtypes in the natural history of HIV in West Africa.

REFERENCES


Address reprint requests to:
Phyllis Kanki
Department of Immunology and Infectious Diseases
Harvard School of Public Health
651 Huntington Ave. FXX-405
Boston, Massachusetts 02115

E-mail: pkanki@hsph.harvard.edu