Sequence Note

HIV Type 1 Diversity and Antiretroviral Drug Resistance Mutations in Burundi

NICOLE VIDAL,1 THEODORE NIYONGABO,2 JEAN NDUWIMANA,3 CHRISTELLE BUTEL,1 ATHANASE NDAYIRAGIJE,2 JOSEPH WAKANA,3 MARTIN NDUWIMANA,2 ERIC DELAPORTE,1 and MARTINE PEETERS 1

ABSTRACT

In 2002, an HIV surveillance study was performed among more than 5500 individuals representing the general population of urban and rural districts in Burundi. In this report, we genetically characterized a subset of the HIV-1-positive samples identified during this survey, including all the HIV-positive samples from Bujumbura, the capital city, and samples from one semiurban and one rural district. One hundred and nineteen samples were genetically characterized in the V3–V5 region of the env gene and/or in the protease and reverse transcriptase region of the pol gene. Phylogenetic analysis of 101 env/pol sequences revealed that the HIV-1 epidemic in Burundi was driven by subtype C (81.2%), followed by subtype A (7.9 %) and polC/envA recombinants (5.9%). One major mutation associated with resistance to antiretroviral drugs (ARVs) in the pol gene, as defined by the International AIDS Society Resistance Testing-USA panel, was observed in one individual, but many minor resistance-associated mutations were also present in the majority of the samples.

The classification of HIV-1 strains in groups (M, N, O), subtypes (A–D, F–H, J, and K), and circulating recombinant forms (CRFs) has helped in tracking the course of the HIV pandemic and allowed us to draw a broad picture of the distribution of HIV strains.1,2 The overall distribution of HIV-1 variants is very heterogeneous, and the highest degree of genetic diversity has been observed in Africa where all HIV-1 groups, subtypes, subsubtypes, and many of the known CRFs circulate.3 Subtypes A, C, and CRF02 predominate, but similar to the global distribution of HIV-1 variants, their geographic distribution on the African continent is heterogeneous and differs from country to country and even within certain countries.4 CRF02_AG predominates in West and West Central Africa, representing 60–90% of HIV-1 variants; in southern Africa subtype C represents more than 95% of the circulating strains and in Eastern Africa, different proportions of A, C, and D cocirculate.5

It is not clear whether differences exist among the various HIV-1 variants (groups, subtypes, and CRFs) in terms of transmissibility, pathogenicity, and responses to antiretroviral therapy. Some differences in disease progression have been seen in Uganda and Tanzania, where subtype D seems to be more pathogenic than subtype A.5,6 Although not associated with higher pathogenicity, higher RNA levels are detected in genital secretions in subtype C-infected patients when compared with subtype A- and D-infected people.7 Some HIV variants are naturally resistant to some classes of antiretroviral drugs, and subtype C strains develop faster resistance to nevirapine than subtype D, which in turn also develops faster resistance than subtype A, in women who received a single dose of nevirapine to prevent HIV transmission from mother to child.8,9 Subtype/CRF distribution is a dynamic process and the genetic diversity of HIV-1 continues to increase and thus remains a permanent challenge to treatment and/or vaccine strategies.4 Continuous monitoring of HIV-1 variants is therefore necessary to adapt treatment and vaccine strategies so that they are efficient against local and contemporary circulating HIV-1 strains.

1UMR145, Institut de Recherche pour le Développement (IRD) and University of Montpellier 1, Montpellier, France.
2Centre de Formation et de Recherche en Médecine et Maladies Infectieuses (CEFORMI)/CHU de Bujumbura, Bujumbura, Burundi.
3Comité National de Lutte contre le SIDA (CNLS), Bujumbura, Burundi.
FIG. 1. Phylogenetic relationships of 456 unambiguously aligned nucleotide sequences representing the V3–V5 region of the env gene (a) and of 1625 unambiguously aligned nucleotide sequences from the protease + reverse transcriptase in the pol gene (b) of the burundian strains. The alignment and the phylogenetic analysis were done as indicated in the text, and HIV-1 group O (HIV-1-ANT70) was used as an outgroup in the neighbor-joining trees. The sequences from Burundi are in black and reference sequences are in gray. The scale bar represents nucleotide substitutions per site. The reference sequences used for the phylogenetic trees in the figure are A1_KRQ2317, A1_UG92UG037, A1_SESOSE7253, A2_97CDKTB48, A2-94CY017, A2-ZMZAM184, A3-DDJ360, A3DDJ579, A3DDJ369, B_FRHXB2R, B_USJRFL, B_USWEAU160, C_ETETH2220, C_IN21068, C_BW96BW0502, D_UG94UG1114, D_ZRNDRK, D_99TCMN011, D_99TCMN012, D91BU009W(ENV), D99GA07412 (POL), D-KE-MB2059(POL), D-01TZA280, (POL), F1_BR93BR020, F1_FILIN9363, F1_BEVI850, F2-96CMMP255, F2-CMMP257, F2-CMMP257, G_BE_DRCBL, G_SESE6165, G_FILH9793, H_BEVI991, H_BEVI997, H_CF90CF056, J_SESE92809, J_SESE91733, K_96CMMP535, K_97CDEQTB11, CRF10-BF071, CRF10-BF061, and CRF10-BF110. The following additional subtype C sequences were added in the env tree: C-ETMCGPI, C-DJ373, C-98TZ017, C-UG268, C-91BU0112, C-97GA71, 97SN1186, C-91BU05, C-96KEMMI14999, C-98BR004, C-99MMR67, C-00GAB20N, C-98SN112HALD, C-96ZM751, C-96BW96BW0502, C-MW5145, C-KER2010, and C-MLDU101. The following additional subtype C sequences were added in the pol tree: C-97TZ04, C-99ZATM10, C-00KE-KER2010, C-01TZBD9, C-ZM-DU151, C-99ZACM9, C-00BW21283, C-01TZA246, C-03SN965HALD, C-99ET-MP1315, and C-00DIMP1482. Only bootstrap values above 75% are indicated.
Burundi is a small country of approximately 6.5 million inhabitants located in East Central Africa. It is bordered by Rwanda in the North, by Tanzania in the South and the East, and by the Democratic Republic of Congo (DRC) in the West, separated from it by the Tanganyka lake. High HIV prevalences have been reported from Burundi, ranging from 3% to 9% in pregnant women from rural and urban areas respectively, but only limited data are available on HIV variants. In 2002, an HIV surveillance study was performed among the general population and a subset of the HIV-1-positive samples identified during this survey were genetically characterized in this report in env and pol, to identify the circulating HIV-1 variants among the general population and to study to what extent antiretroviral (ARV) drug resistance mutations circulate.

The HIV serosurveillance study was conducted among the general population (>12 years old), using an anonymous non-correlated cluster sampling method. A total of 30 randomly selected clusters of people were investigated from Bujumbura, the capital city, and 30 randomly selected clusters from semi-urban and rural areas covering 13 provinces. A total of 5569

FIG. 1. (Continued).
individuals were tested for HIV antibodies, 1053 from Bujumbura, 1062 from semiurban, and 3454 from rural areas. Blood samples as well as social and demographic information (sex, age, marital status, number of children, profession, and education) were collected anonymously according to the recommendations of the National Authorities. Whole blood was collected on EDTA tubes and buffy coats and plasmas were separated and stored at −20°C. Buffy coats were shipped to IRD (Montpellier, France) on dry ice for further genetic characterization. The plasma samples were screened locally for the presence of HIV antibodies by ELISA (Genscreen, BioRad). ELISA-reactive samples were further tested with four rapid tests [Determine (Abbott), Genie II (Bio-Rad), Capillus (Trinity Biotech), and HIVspot (Genelabs)]. Samples with discordant serological results were tested with the InnoLia HIV confirmation test (Innogenetics, Belgium). Among the HIV-1-positive samples identified in the survey, a subset of 119 was genetically characterized in env and/or pol and included 96 of the 104 HIV-1-positive samples that were identified in the capital city, Bujumbura, 11 of 15 HIV-1 positives from Gitega, a semiurban area in the middle of the country, and 12 of the 15 HIV-1 positives that were identified in rural areas from the Bujumbura province. Overall, the median age of the samples included in the HIV-1 genetic study was 36.4 years (from 12 to 70 years old) and 59% of the samples were from women and 41% from men, also reflecting the overall higher HIV prevalence rates among women in the general population from Burundi.

DNA was extracted from peripheral blood mononuclear cells (PBMCs) present in buffy coats by using the Qiagen DNA isolation kit (Qiagen SA, Courtaboeuf, France) according to the manufacturer’s instructions. The genetic subtypes were identified in env and pol as previously described. Briefly, the envelope outer primers were env1 and env2 and previously described ES7 and ES8 were used as inner and sequencing primers, generating a fragment of approximately 670 base pairs. For pol, a fragment of approximately 1865 base pairs, corresponding to the protease and reverse transcriptase region, was amplified by a nested PCR using Expand High Fidelity PCR system (Roche, France) with outer primers G25REV and IN3 and inner primers AV150 and polM4. The amplified env and pol fragments were purified (GeneClean Turbo kit, Q-Biogen, Illkirch, France) and directly sequenced using Big Dye Terminator v3.1 (Applied Biosystems, France) according to the manufacturer’s instructions. Electrophoresis and data collection were performed on an Applied Biosystems 3100 genetic analyzer.

The newly determined sequences were aligned with known representatives of the different subtypes, subsubtypes, and CRFs described in West and Central Africa using CLUSTAL W with minor manual adjustments, bearing in mind the protein sequences. Regions that could not be aligned unambiguously, due to length or sequence variability, were omitted from the analysis. Phylogenetic trees using the neighbor-joining method and reliability of the branching orders using the bootstrap approach were implemented with CLUSTAL W. Genetic distances were calculated with the Kimura’s two parameter method. In addition to the pure subtypes we added the alignment reference strains representing the following variants that have been documented in Africa: CRF01_AE, CRF02_AG, CRF04_cpx, CRF05_DF, CRF06_cpx, CRF09_cpx, CRF10_CD, CRF11_cpx, CRF13_cpx, CRF18_cpx, CRF19_cpx, subtype A2, and the preliminary characterized subtype L. To clearly identify whether a sequence belonged to a subgroup corresponding to a certain subtype or CRF, phylogenetic analysis was done for each sequence individually and subtype/CRFs designations were assigned when bootstrap values were above 80%. Then different trees were constructed for each group of new sequences that were thought to cluster together and, finally, a general tree was obtained to visualize the results. The clustering of each new sequence was compared and should be concordant between all trees. For better clarity, the general trees in Fig. 1 were drawn with the minimal number of references, i.e., without some CRF and unique sequences when not represented among the samples from Burundi. The new pol sequences were further investigated by bootscan and similarity analyses using the Simplot software to determine whether they were pure subtypes or CRFs. The new V3–V5 sequences were not analyzed by bootscan and similarity plot analysis because the length of the sequenced fragment was not sufficient to perform reliable analyses. Similarity plots, using Simplot 2.5 Software (S. Ray, http://www.med.jhu.edu/deptmed/stay), determined the percent similarity between a new sequence and selected groups of reference sequences, by moving a window of 400 base pairs with 20 base pair increments along the alignment. Similarity values were plotted at the midpoint of the 400-base pair fragments. The Simplot 2.5 software was also used to generate bootscan plots: bootscanning was performed on neighbor-joining trees by using Seqboot, DNAdist (with Kimura’s parameter method and a transition/transversion ratio of 2.0), Neighbor and Consense from the PHYLIP package, for a 400-base pair window moved along the alignment in increments of 20 base pairs. One thousand bootstraps replicates were evaluated for each phylogeny, and the bootstrap values for the studied sequences were plotted at the midpoint of each window. In these two sets of analyses, the new sequences were aligned with consensus sequences (50% threshold) representing the nonrecombinant subtypes from the same alignment used for the phylogenetic analyses.

Table 1 shows the detailed distribution of the HIV-1 strains circulating in Burundi. In Bujumbura, 84 samples were genetically characterized in env and pol, 10 in env only, and 2 in pol only. Similarly, in semiurban Gitega and rural Bujumbura, 7 and 10 samples were respectively characterized in env and pol, then 4 and 1 in env only, and 0 and 1 in pol only. Overall, subtype C was largely predominant in the country and represented 82% (96/116) of the env sequences and 87% of the pol sequences, respectively. Among the remaining samples, subtype A predominated and represented 13.8% (16/116) of the env sequences, and 8.6% (9/104) of the pol sequences. Two subtype D samples, one CFR10 and other minor variants, were also identified, one env subtype H and one pol G/J mosaic.

On the total of 101 samples that were simultaneously characterized in the env and pol regions, nine had different subtype designations among both genes: six envA/polC, one envA/pol G/J, one envH/polA, and one env A strain revealed to be a complex mosaic virus in pol involving subtypes A and C and a fragment that could not be classified. Overall, about 9% of the HIV-1 strains circulating in Burundi are thus recombinants. Interestingly, all the env subtype C samples were also subtype C in pol, however, only half of env subtype A samples were subtype A in pol.

Figure 1 illustrates the phylogenetic trees obtained in the V3–V5 region of the env gene (Fig. 1a) and in the protease and reverse transcriptase regions of the pol gene (Fig. 1b). In the phylogenetic tree analysis, we added various previously described subtype C sequences from different geographic regions.
HIV-1 DIVERSITY IN BURUNDI

TABLE 1. NUMBERS OBSERVED FOR EACH COMBINATION OF env/pol SUBTYPE/CRF DESIGNATIONS FROM BUJUMBURA

<table>
<thead>
<tr>
<th>env/pol</th>
<th>Bujumbura</th>
<th>Gitega</th>
<th>Rural Bujumbura</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/A</td>
<td>7</td>
<td>—</td>
<td>—</td>
<td>7</td>
</tr>
<tr>
<td>A/C</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>A/G-J</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>A/A-C-U</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>A/nt</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>C/C</td>
<td>67</td>
<td>6</td>
<td>9</td>
<td>82</td>
</tr>
<tr>
<td>C/nt</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>nt/C</td>
<td>2</td>
<td>—</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>D/D</td>
<td>2</td>
<td>—</td>
<td>—</td>
<td>2</td>
</tr>
<tr>
<td>H/A</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>CRF10/CRF10</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>96</td>
<td>11</td>
<td>12</td>
<td>119</td>
</tr>
</tbody>
</table>

Samples were from Bujumbura, the capital city; Gitega, a semiurban region; and from rural Bujumbura. The subtypes/CRFs were determined by phylogenetic tree analysis based on phylogenetic tree and bootscan analyses.

A/nt sequence was not available due to negative PCR or limited sample volume.

in Africa (Ethiopia, Djibouti, Somalia, Tanzania, Uganda, Burundi, Kenya, Gabon, Senegal, Mauritania, Zambia, Bostwana, Malawi, and South Africa). No particular geographic clustering was observed between the subtype C strains from Burundi and those from certain geographic areas. It was not possible to include the same references for subtype C isolates from different geographic origins in the env and pol trees, because they were not always available for both gene fragments. The intrasubtype genetic distance calculated on 85 subtype C strains from Burundi over 1625 unambiguously aligned base pairs in the pol gene is 5.34%, and is comparable to the intrasubtype genetic distance among the 14 C reference sequences from different geographic locations (5.19%). Similarly, the intrasubtype genetic distance calculated on 95 subtype C strains from Burundi over 456 unambiguously aligned base pairs in the env gene is 13.38%, comparable also to that of the 22 subtype C reference sequences (11.68%).

Sequences were also analyzed for the presence of major and minor mutations in protease and reverse transcriptase genes at positions known to be associated with drug resistance according to the last update (October–November, 2005) from the International AIDS Society Resistance Testing USA panel (http://www.iasusa.org/resistance_mutations/index.html) and by three drug resistance interpretation algorithms (ANRS 2005.7, Stanford Database version 4.1.7, Rega version 6.4.1), using a Stanford Resistance database-tool, HIVValg version 4.1.7 (http://hivdb6.stanford.edu/asi/deployed/hiv_central.pl?program=hivalg&action=showSequenceForm). Analysis of the reverse transcriptase (RT) sequences revealed one sample from Bujumbura with the G190E mutation conferring resistance to NNRTIs (nonnucleosidic inhibitors of the RT) according to the three algorithms but which is not mentioned in the IAS list. Five other samples were identified with each one a minor resistance-associated mutation to certain NNRTIs: E44D (n = 2), V118I (n = 1), T69S (n = 1), and T215A (n = 1). The single presence of T69S did not lead to the prediction of decreased susceptibility by the algorithms and the T215A mutation is not equally considered in drug resistance interpretation algorithms and is not mentioned at all in the IAS list. Polymorphism at a position associated with major resistance mutations in RT was seen in one sample only (K103R). No major mutations were seen in the protease gene, however, many minor mutations characteristic for non-B HIV-1 variants were observed. The number and combination of some of these mutations are interpreted as associated with intermediate or low-level resistance to some protease inhibitors (PIs) by certain drug resistance interpretation algorithms for seven patients. Overall, the following mutations associated with major resistance were present in the strains from Burundi: L10(FV) (n = 8, 7.6%), L10M (n = 2, 1.9%), K20(R/I) (n = 17, 16.2%), L33V (n = 2, 1.9%), M36I (n = 94, 89.5%), M36 (T/L/V) (n = 3, 2.9%), D60E (n = 8, 7.6%), L63P (n = 22, 21%), L63(A/H/S/T/V) (n = 26, 24.8%), A71T (n = 1), and I93L (n = 91, 86.7%). The most represented substitutions were M36I (89.5%) and I93L (86.7%); all have been previously shown to occur as natural variants in non-B HIV-1 strains with I93L specific for subtype C.8,17 In addition, polymorphisms at positions of major mutations were noted in the protease gene: G48E (n = 1), V82I (n = 17, 16.3%), and L90 (L/F) (n = 1).

In conclusion, we have shown that the HIV-1 epidemic in Burundi is principally driven by strains of subtype C, but A and A/C recombinants are also identified. This subtype distribution is in agreement with what has previously been reported on a limited number of samples from Burundi.11 but seems to be relatively different from the neighboring countries. In Tanzania subtypes A, C, and D cocirculate at almost equal levels with many unique recombinants involving two or three of the cocirculating subtypes.18,19 In Rwanda, the epidemic seems to be driven by subtypes A and C, but C strains were significantly less prevalent than A.20 In Uganda, the HIV-1 epidemic is driven by subtypes A and D, with unique A/D recombinants emerging, but subtype C and even G have also been noted.21,22 In Kenya, subtype A predominates, but D and C cocirculate also.23 In the Democratic Republic of Congo, subtype C is only predominant in the southern part of the country whereas in the other parts of this country an extremely high genetic diversity is seen in a number of cocirculating HIV-1 variants.4

We have also shown that rare resistance mutations to the RT inhibitors can be found in the general population. However, the mutations were not equally considered by the different algorithms, similar to what has previously been reported by other studies evaluating the performance of drug resistance interpretation algorithms on non-B HIV-1 variants.24,25 In this study we combined the efforts from the sentinel serosurveillance studies to analyze the HIV-1 subtype/CRF distribution and the eventual presence of drug resistance mutations. However, by this approach, the proportion of resistant strains in the general population does not directly reflect recently transmitted resistance, because we do not know whether the HIV-positive individuals are completely ARV naive. It will be necessary to continue the surveillance of circulating HIV variants. Now that access to HIV treatment is expanded, surveillance of drug-resistant strains over time should be implemented also, but by preference on population groups that represent individuals who were recently infected with HIV as recommended by WHO or
for whom data on previous ARV treatment can be collected. The data from these programs will contribute significantly to the evaluation of the efficacy of regimens and ARV programs as well as to HIV-1 subtype/CRF distribution, and provide important public health information.

ACCESSION NUMBERS

Sequences have been submitted to EMBL under the following accession numbers: from AM260218 to AM260322 and AM260323 to AM260437 for the pol and env sequences, respectively.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the Ministry of Health, the National AIDS/STD programme, and the National Ethical Committee for permission to perform this survey. The study was sponsored by the World Bank (HIV surveillance study) and IRD.

REFERENCES


Address reprint requests to: Martine Peeters
Laboratoire Retrovirus
UMR145, IRD
911 Avenue Agropolis
BP 64501
34394 Montpellier Cdx 1, France
E-mail: martine.peeters@mpl.ird.fr
This article has been cited by:


