Volume 18, Number 5, 2002, pp. 313–325

Molecular Epidemiology of HIV Type 1 Infection in Portugal: High Prevalence of Non-B Subtypes

AIDA ESTEVES,[1] RICARDO PARREIRA,[1] TERESA VENENNO,[1] MARGARIDA FRANCO,[2] JOÃO PIEDADE,[1] JOSÉ GERMANO DE SOUSA,[2] and WANDA F. CANAS-FERREIRA[1]

ABSTRACT

In this study, we have investigated the diversity of current HIV-1 strains circulating in the metropolitan area of Lisbon, Portugal. A total of 217 HIV-1-positive blood samples, collected between October 1998 and December 2000, was genetically characterized in the gp120 C2V3C3 region ($n = 205$) or part of the gp41 N-terminal segment ($n = 12$) by heteroduplex mobility assay (HMA) and/or DNA sequencing. The HMA subtyping efficiency (number of samples unambiguously subtyped by HMA divided by the total number of samples subtyped) was 65.9% (143 of 217), with indeterminate migration patterns of subtype A and G strains contributing significantly to this value. On the overall, subtype B was the most prevalent (50.2%), followed by subtypes G (21.7%), A (17.5%), and F (5.5%), whereas subtypes C, D, H, and J accounted altogether for 5.1% of the infections. Non-B subtypes were responsible for 77.4 and 33.1% of the infections among African immigrants and Portuguese subjects, respectively. Angolan individuals ($n = 25$) were the only ones infected with all the HIV-1 subtypes documented, probably reflecting a high degree of viral genetic diversification in their country of origin. Phylogenetic analysis showed a predominance of IbNG-like viruses among subtype A sequences and two new major subclusters within subtype G (G^P and $G^{P\prime}$). The majority of the Portuguese G sequences described formed a well-defined subcluster (G^P), supported by bootstrap values >90%, phylogenetically distant from clade G sequences in databases. *gag* (p24/p7) sequence analysis of these variants confirmed the maintenance of the subtype G subclusters. The multiple subclustering observed for the major clades A, B, D, and G, as well as the variety of subtypes found, indicate a high diversity of HIV-1 variants circulating in Portugal and suggest a need for continuous epidemiologic surveillance.

INTRODUCTION

THE FIRST CASE OF HIV-1 infection in Portugal was described in 1983, at about the same time the virus was reported in other Western European countries and the United States. Since 1991, the HIV seroprevalence rates have increased rapidly. By the end of 1999, the estimated number of people, aged 15–49 years, living with HIV/AIDS in Portugal was 36,000, corresponding to an adult prevalence of 0.74%, the highest value in Western Europe.[1] After an initial period dominated by homosexual transmission, a shift toward transmission through heterosexual contacts and drug injection was noticed. Nowadays, intravenous drug users (IVDUs) are the main risk group in the epidemics and account for 58% of the AIDS cases diagnosed in 1999.[2]

One of the most outstanding properties of HIV is its unusually high genetic variability. HIV-1 sequence data accumulated in the last two decades revealed three distinct lineages: M (main), O (outlier), and N (new or non-M, non-O). Viruses from group M dominate the pandemic and include 9 major clades (A–D, F–H, J, and K) and 12 recombinant circulating forms (CRFs) (HIV Sequence Database, http://hiv-web.lanl.gov/). Within some subtypes further phylogenetic structure can be identified, leading to a classification into subsubtypes.[3] Despite the increasing amount of data attempting to assign distinct biological properties to individual HIV subtypes,[4] differences in transmission and disease progression rates and in antiretroviral susceptibility, as well as its importance for vaccine design, still need clarification. However, phylogenetic analysis of circulating viruses has proved useful in tracking the spread of HIV in

[1]Unidade de Virologia/Unidade de Parasitologia e Microbiologia Médicas, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, P-1349-008 Lisbon, Portugal.
[2]Serviço de Patologia Clínica, Hospital Fernando da Fonseca, P-2700-276 Amadora, Portugal.

different populations and geographic regions. Therefore, this information is fundamental in defining strategies for the prevention and control of infection as well as for the implementation of proper diagnostic tests and viral load assays.

The greatest diversity of HIV-1 strains was demonstrated in sub-Saharan Africa, where all the subtypes have been found.[4] In Europe, HIV-1 infection is currently dominated by subtype B but a growing mixture of HIV-1 subtypes has been reported. Studies in France,[5,6] Sweden,[7] Denmark,[8] Switzerland,[9] the United Kingdom,[10] Spain,[11,12] Greece,[13] Cyprus,[14] Belgium,[15,16] The Netherlands,[16] Germany,[17] Italy,[18] Austria,[19] Norway,[20] and Eastern Europe[21–27] have detected most of the group M clades. Group O has been reported in France,[6] Germany,[28] Belgium,[29] and Spain.[30] Most of the non-B HIV-1 subtypes described for European countries originated from African immigrants and European natives who acquired these strains abroad or by contact with immigrants.

Portugal has strong historic links and maintains ongoing relationships with several (mainly West) African countries endemic for HIV. The potential for introduction of multiple HIV-1 subtypes in Portugal resides not only in the high numbers of immigrants from those regions who enter the country, but also in those Portuguese natives returning home after living and working in Africa. The AIDS epidemic in Portugal is also characterized by a relatively high number of HIV-2 infections.[2] Phylogenetic analysis of HIV-2 from infected individuals residing in Portugal showed that subtype A is responsible for the infections studied.[31] However, subtyping data are not available for HIV-1 strains. Because no prior work has documented the molecular epidemiology of the Portuguese HIV-1 epidemic, the present study was undertaken to survey circulating subtypes and transmission patterns of viruses from seropositive individuals, both persons of Portuguese origin and African immigrants, living in the urban area of Lisbon.

MATERIALS AND METHODS

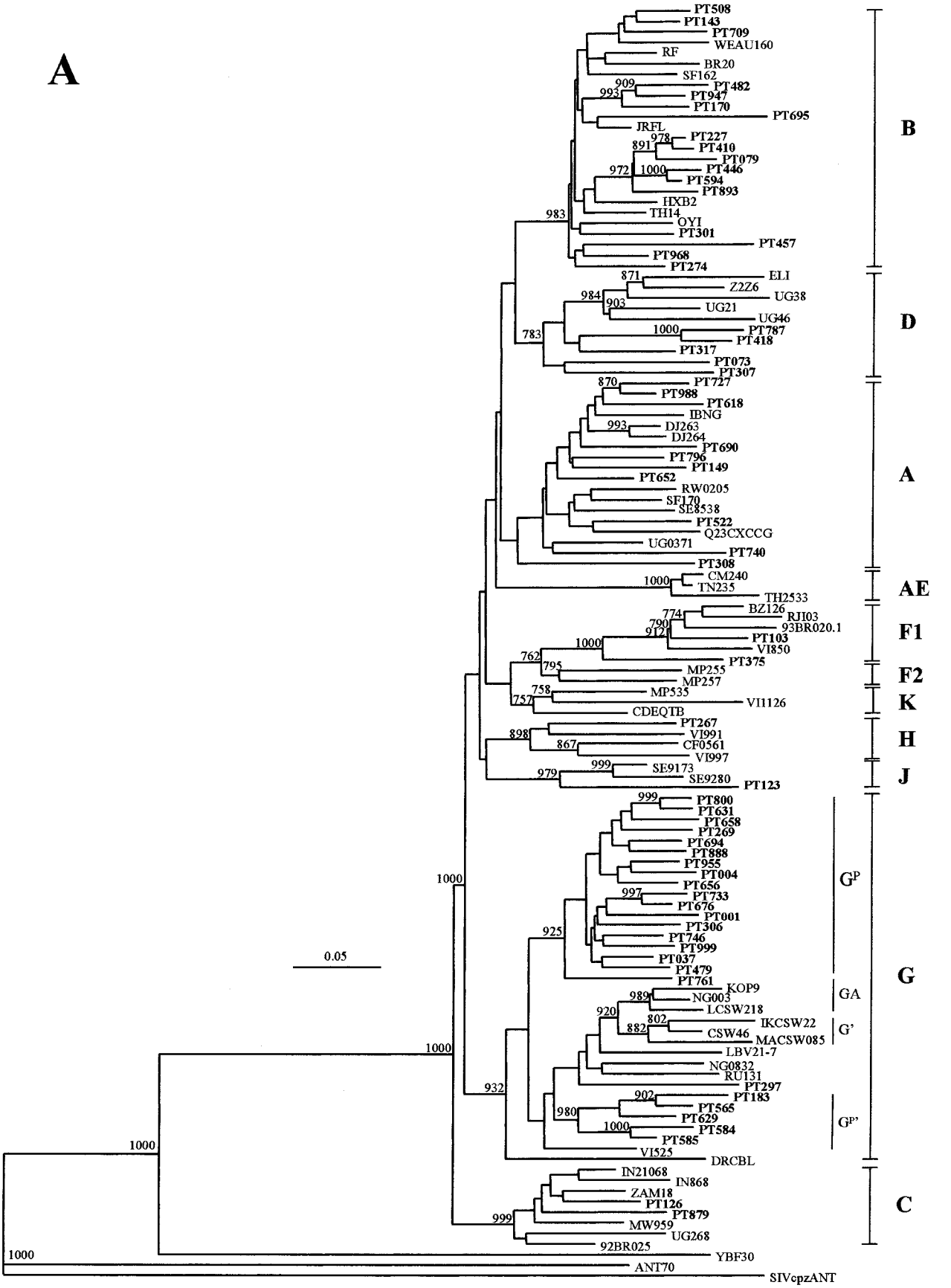
Subjects

In 1998, a collaboration protocol for HIV-1 subtyping was established between Hospital Fernando da Fonseca (HFF, Amadora, Portugal) and our laboratory. HFF provided us with 5-ml blood samples, collected with informed consent, as well as with anonymous epidemiological data, including age, sex, citizenship, and most likely route of infection. HFF is localized in the Lisbon metropolitan area and serves a heterogeneous population of white Portuguese citizens and several generations of individuals originating from former Portuguese African colonies. From October 1998 to December 2000, 241 blood specimens from HIV-1-seropositive individuals (102 females and 139 males), aged 16 to 73 years (average, 35.4 years), were received for analysis. The study group included 141 individuals of Portuguese nationality and 92 Africans (8 with nonreported origin), all residents of Portugal. The main self-reported modes of infection were intravenous drug use ($n = 71$) and heterosexual transmission ($n = 109$). Six subjects were coinfected with HIV-2. The spectrum of $CD4^+$ lymphocyte cell counts ranged from <10 to 1167 cells/μl; 25% of patients had counts of less than 200 cells/μl, indicating that many patients had advanced disease.

Polymerase chain reaction and heteroduplex mobility assay

Blood samples were collected in 3.8% sodium citrate and peripheral blood mononuclear cells (PBMCs) were separated by Lymphoprep (Nycomed, Askes, Norway) density centrifugation. Total cellular DNA was prepared by lysis of 10^7 PBMCs/ml in lysis buffer (10 mM Tris-HCl [pH 8.3], 0.45% Nonidet P-40 [NP-40], 0.45% Tween 20), addition of proteinase K (10 mg/ml), incubation at 56°C for 1 hr, and inactivation of proteinase K at 95°C for 20 min. *env* C2V3C3 proviral sequences were amplified from 5–10 μl of cell lysate by nested polymerase chain reaction (PCR), according to the recommendations included with the NIH AIDS Research and Reference Reagent Program (Bethesda, MD) heteroduplex mobility assay (HMA) kit. Primer sets ED3/ED14 and ED31/ED33 were routinely used for the first and second rounds of amplification, respectively. Whenever amplification failed to yield a product visible on ethidium bromide–agarose gels, one or more of the following approaches were used: (1) alternative second-round primers ED5/ED12 or ES7/ES8; (2) five initial cycles at lower annealing temperature (42–50°C); (3) triple-nested PCR with primer set ED3/ED14 followed by primer sets ED5/ED12 and ED31/ED33. Samples that still could not be amplified by use of HMA standard primers were tested with the generic primer sets gp40F1/gp41R1 and gp46F2/gp47R2, designed to detect all groups of HIV-1, according to Yang *et al.*[32] PCR for a cell gene (β_2-microglobulin) was performed as described elsewhere,[33] for all the specimens in which every attempt of HIV-1 sequence amplification failed. To avoid cross-contamination, physical confinement measures and the use of aerosol-resistant tips were applied to the preparation of PCR mixtures, template extraction, and manipulation of PCR products, for both subject samples and references. Negative con-

FIG. 1. (**A**) Phylogenetic relationships between HIV-1 strains from Portuguese samples, and representative HIV-1 strains from groups M, subtypes A to J (as indicated), O (ANT70), and N (YBF30), using SIVcpzANT as the outgroup sequence. The phylogenetic tree (neighbor-joining) was created from distance matrixes calculated by the Kimura two-parameter method, based on 502 unambiguously aligned nucleotides covering the C2V3C3 genome region. The DNA sequences described in this report are indicated in boldface type. In the G clade, G′ and GA (GA-NG003) are the subgroups defined by Peeters *et al.*[41] Branch nodes supported by bootstrap values above 750 (of 1000 resamplings) are indicated. The scale bar represents 5% of genetic diversity. (**B**) Neighbor-joining tree based on *env* gp41 nucleotide coding sequences (372 nucleotides) from representative group M (as indicated) and O (MVP5180) sequences, with SIVcpzUS as the outgroup. Samples newly identified in this study are indicated in boldface type. Values at the branches represent bootstrap values exceeding 750 of 1000 resampling. The scale bar represents the evolutionary distance of 0.1 nucleotides per position in the sequence.


A
PT508
PT143
PT709
WEAU160
RF
BR20
SF162
909
PT482
993
PT947
PT170
PT695
JRFL
978
PT227
891
PT410
PT079
972
1000
PT446
PT594
PT893
HXB2
TH14
983
OYI
PT301
PT457
PT968
PT274
B
871
ELI
Z2Z6
984
UG38
903
UG21
UG46
1000
PT787
783
PT418
PT317
PT073
PT307
D
870
PT727
PT988
PT618
IBNG
993
DJ263
DJ264
PT690
PT796
PT149
PT652
RW0205
SF170
SE8538
PT522
Q23CXCCG
UG0371
PT740
PT308
A
1000
CM240
TN235
TH2533
AE
774
BZ126
RJI03
790
93BR020.1
912
PT103
1000
VI850
762
PT375
F1
795
MP255
MP257
F2
758
MP535
757
VI1126
K
CDEQTB
PT267
VI991
898
CF0561
867
VI997
H
999
SE9173
979
SE9280
PT123
J
999
PT800
PT631
PT658
PT269
PT694
PT888
PT955
1000
PT004
PT656
997
PT733
PT676
PT001
925
PT306
PT746
PT999
0.05
PT037
PT479
PT761
GP
G
KOP9
989
NG003
GA
LCSW218
920
802
IKCSW22
882
CSW46
MACSW085
G'
1000
LBV21-7
NG0832
RU131
PT297
932
902
PT183
PT565
980
PT629
GP'
1000
PT584
PT585
VI525
DRCBL
1000
IN21068
IN868
ZAM18
PT126
PT879
C
999
MW959
UG268
92BR025
1000
YBF30
ANT70
SIVcpzANT

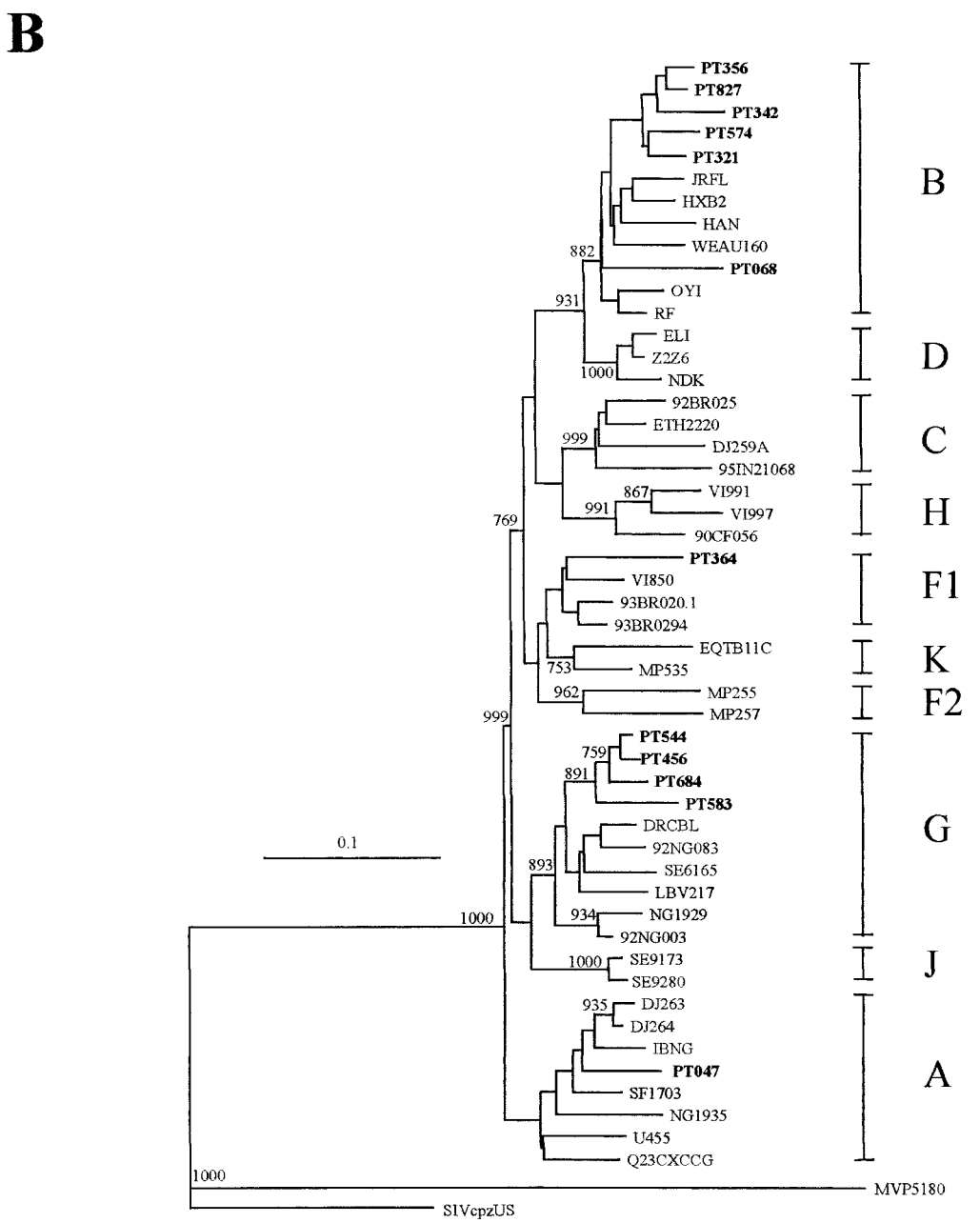


FIG. 1. *Continued.*

trols were included in all PCRs. The gp120-coding region PCR products from study subjects were genetically compared with subtype reference sequences by heteroduplex mobility assay as described by Delwart *et al.*[34] Subtype references routinely used were as follows: A2 (IC144), B3 (SF162), C2 (IN868), D1 (UG21), E2 (TH06), F1 (BZ162), G1 (LBV217), G2 (VI525), and H2 (VI557).

For genetic analysis of *gag* sequences, part of the p24/p7-coding region of the HIV-1 genome was amplified with primer pairs H1G777/H1P202 and H1gag1584/g17, using reaction conditions described for *gag* HMA by Heyndrickx *et al.*[35]

DNA sequencing and phylogenetic analysis

Samples exhibiting an indeterminate HMA pattern or an electrophoretic mobility (R_f) below 0.4 compared with the homoduplexes were subtyped by phylogenetic analysis of *env* (C2V3C3 and gp41). The HIV-1 C2V3C3 DNA fragments analyzed corresponded to PCR products (approximately 500 bp) amplified with primers ED31/ED33. gp41 or *gag* nucleotide sequences were obtained from PCR fragments amplified by primer sets gp46F2/gp47R2 and H1gag1584/g17, respectively. The different amplicons were directly cloned into pGEM-T Easy (Promega, Madison, WI) and individual clones were se-

quenced with fluorescent dye-coupled dideoxynucleotides in an ABI PRISM 377 automated sequencer (Applied Biosystems, Foster City, CA).

DNA and predicted protein sequences were aligned with CLUSTAL W and manually stripped for global gaps. Phylogenetic analysis was performed with both PHYLIP version 3.5 and PUZZLE version 4.0.2. software packages (available online at http://www.igc.gulbenkian.pt). Phylogenetic trees were constructed by the neighbor-joining method (NEIGHBOR), from pairwise distance matrixes calculated with the Kimura two-parameter model (using DNADIST), or by the maximum likelihood approach (implemented in the DNAML program), based on transition–transversion ratios calculated from the input data with the HKY85 model of substitution.[36] The robustness of the generated trees was evaluated by bootstrap analysis of 1000 resamplings, using SEQBOOT. CONSENSUS was used to determine the consensus tree, which was then drawn with TREEVIEW.[37] The putative mosaic structure of nucleotide sequences was investigated with RIP (Recombinant Identification Program). PCOORD (Principal Coordinate Analysis) and SUDI (Subtyping Distance Tool) were used as tools for the definition of subsubtypes. For the latter, we used the following reference sequences: subtype A (Q23CXCCG, SOSE7253, TZSE8538, UGSE6594, UGSE8891, 92UG037, U455, and UG273A); subtype B (HAN, CAM1, JH32, MN, NY5CG, P896, RF, SF2, and YU2); subtype C (92BR025, 96-BW0502, CJ259A, ETH2220, 301904, 301905, 301999, and UG266A2); subtype D (SE365A2, 94UG1141, UG266A2, UG274A2, 84ZR085, JY1, NDK, and Z2Z6); subsubtype F1 (93BR0201, BZ126A, BZ163A, and FIN9363); subsubtype F2 (MP255 and MP257); subtype G (DRCBL, HH8793, LBV217, 92NG083, and SE6165); subtype H (90CF056, VI991, and VI997); subtype J (SE9173 and SE9280); subtype K (EQTB11C and MP535); and the outgroup sequence SIVcpzGAB1. The genetic distances between subtypes B and D, as well as subsubtypes F1 and F2, were used as intersubsubtype references. HIV-BLAST was used to search for HIV DNA sequences with the highest homology to those presented in this study. RIP, PCOORD, SUDI, HIV-BLAST, and also all the reference sequences used, are available online at the HIV Sequence Database (http://hiv-web.lanl.gov/).

Statistical analysis

Statistical analysis included, when appropriate, p values of the χ^2 test, the χ^2 test with Yates' correction, and the two-tailed Fisher exact test.

RESULTS

Subtyping and epidemiological data

During the 2-year study period, proviral DNA extracted from 241 blood samples was amplified in the *env* C2V3C3 region by nested PCR. No amplification product was obtained in 36 cases (14.9%) despite repeated attempts, alternative PCR protocols, and successful amplification of β_2-microglobulin DNA. Two-thirds of the nonamplified samples were derived from both Portuguese and African individuals with $CD4^+$ counts >500 cells/μl and/or viral loads $<10^4$ copies/ml; this could, at least in part, account for these results. These samples were subjected to PCR amplification with gp41 primers, which amplify HIV-1 isolates of groups M, N, and O,[32] and proviral DNA was successfully amplified in 12 of them (33.3%). Of the 12 samples whose DNA was amplified with the gp41 primers, but not with the C2V3C3 primers, 7 corresponded to individuals with high

Table 1. B and Non-B HIV-1 Subtype Distribution among Residents in Portugal (Lisbon Metropolitan Area), October 1998–December 2000

	Total samples subtyped		*Subtype B*		*Non-B subtypes*		
Characteristic	n	%	n	%	n	%	p *Value*[a]
All samples	217	100	109	50.2	108	49.8	
Sex							NS
Male	127	58.5	67	52.8	60	47.2	
Female	90	41.5	42	46.7	48	53.3	
Age group (years)							NS
16–29	78	35.9	43	55.1	35	44.9	
30–44	98	45.2	53	54.1	45	45.9	
>45	39	18.0	13	33.3	26	66.7	
Not stated	2	0.9	0	0	2	100	
Citizenship							<0.001
Portugal	127	58.5	85	66.9	42	33.1	
Sub-Saharan Africa	84	38.7	19	22.6	65	77.4	
Other/not stated[b]	6	2.8	5	83.3	1	16.7	
Transmission risk							NS
IVDU	65	30.0	44	67.7	21	32.3	
Heterosexual contact	95	43.8	39	41.1	56	58.9	
Other/not stated[c]	57	26.3	26	45.6	31	54.4	

[a] χ^2 test or χ^2 test with Yates' correction (NS, not significant for the level of significance tested).
[b] Brazil (2), Spain (1), not stated (3).
[c] Homosexual contact (3), transfusion recipient (2), not stated (52).

$CD4^+$ cell counts and low viral loads, suggesting that some of the previous negative results were not due to a low input of proviral DNA.

Screening of the 205 amplified C2V3C3 sequences by HMA unequivocally subtyped viruses from 143 samples (69.8%), with subtype distribution as follows: 27 A, 86 B, 1 C, 9 F, 19 G, and 1 H. C2V3C3 amplicons from the remaining 62 cases, with dubious results by HMA, were cloned and sequenced. Comparative phylogenetic analysis (Fig. 1A) confirmed 17 B sequences with low R_f values in HMA gels and identified 24 subtype G strains, 10 subtype A (7 of which cluster with CRF02_AG sequences, but only with a bootstrap value of 706), 5 subtype D, 2 each for subtypes F and C, and 1 each for subtypes H and J. It is noteworthy that most of the subtype G samples had an undetermined result by HMA because of a similar electrophoretic migration profile of the heteroduplexes formed with subtype A and G references supplied with the kit. The phylogenetic tree in Fig. 1A shows a clear separation of the different HIV-1 subtypes, the great majority of which are supported by high bootstrap values. The single exception corresponds to the one defining subtype A (bootstrap value of 622). However, all the defined subtypes are in agreement with the results previously obtained by HIV-BLAST. All the nucleotide sequences segregating as separate branches from the main clusters were analyzed by RIP (with a window of 80 nucleotides and a threshold of significance of 90%) to try to define intragenic mosaic structures but no statistically significant results were obtained. A complementary analysis (using maximum likelihood) based on the group of sequences found to belong to subtype B showed that three of them (PT170, PT947, and PT482) cluster with a group of HIV-1 sequences described as GB recombinants[12] with a statistically significant bootstrap value of 917 (data not shown). Although our analysis did not include analysis of the *pol* sequences that characterize the HIV-1 GB recombinants, we tentatively suggest that the three sequences mentioned previously might have been amplified from GB recombinant proviral DNA. Another strongly supported subgroup of B sequences (clustering with a bootstrap value of 972) is evident from the NJ tree analysis, suggesting a possible epidemiological link between sequences PT227, PT410, PT079, PT446, PT594, and PT893.

Phylogenetic analysis performed for gp41 nucleotide sequences (Fig. 1B) revealed six strains belonging to subtype B, one to F, four to G, and one to A. All the subtype G gp41 sequences and five of the six subtype B gp41 sequences formed separate clusters in the tree. Although the clustering of HIV-1 subtype B variants in a singular subgroup is not clearly bootstrap supported (bootstrap value of 689), the same does not apply to the G sequence cluster, which is held separately by a bootstrap value of 891 from any of the subtype G sequences use.

The overall results from the three subtyping approaches are summarized in Tables 1 and 2. Subtype B is present in 109 (50.2%) of the 217 specimens tested. Forty-seven individuals (21.7%) are infected with subtype G, 38 (17.5%) with subtype A, and 12 (5.5%) with subtype F. Subtypes C, D, H, and J account altogether for 5.1% of the infections in the population studied. Analysis of epidemiological and subtyping data (Table 1) demonstrates that there is no association between subtype distribution (B vs. non-B) and sex, age, or transmission (Table 1). However, a strong association between HIV-1 subtype distribution (B vs. non-B) and the geographic origin of the seropositive individuals (Portuguese vs. immigrants from sub-Saharan Africa) (χ^2 test with Yates' correction, $p < 0.001$) is observed. Subtype B was significantly more frequent in Portuguese than in individuals originating from Africa (66.9 vs. 22.6%), whereas non-B strains were found predominantly among Africans (77.4%). Phylogenetic analysis for both C2V3C3 and gp41 sequences (Fig. 1A and B) does not show significant clustering of samples that could be correlated with geographic origin, age, or sex of the subjects from whom they were derived, and the branching pattern of sequences from individuals within risk groups does not suggest founder effects.

Table 2 shows that subtypes G and A are the most abundant non-B subtypes in both Portuguese and immigrant populations, accounting altogether for 78.7% of all non-B viruses. The great majority of the immigrants studied ($n = 84$) originated from sub-Saharan Africa, namely from five different Portuguese-speaking countries. Among HIV-1 strains from Angolan individuals we found eight established *env* subtypes (A, B, C, D, F, G, H, and J), suggesting a high diversity of group M subtypes circulating in that country. Cape Verde Islands and São Tomé and Príncipe samples are both infected with subtypes A, B, F, and G, whereas among those from Guinea-Bissau subtype A is prevalent.

TABLE 2. HIV-1 SUBTYPE DISTRIBUTION BY CITIZENSHIP, AMONG 217 INFECTED INDIVIDUALS LIVING IN PORTUGAL (LISBON METROPOLITAN AREA), OCTOBER 1998–DECEMBER 2000

	HIV-1 subtype								
		Non-B							
	B	*A*	*C*	*D*	*F*	*G*	*H*	*J*	*Total*
Portugal	85	13	1	1	4	23	0	0	127
Angola	2	6	2	2	1	9	2	1	25
Cape Verde	9	8	0	0	5	12	0	0	34
Guinea-Bissau	4	9	0	2	1	1	0	0	17
Other/not stated[a]	9	2	0	0	1	2	0	0	14
Total:	109	38	3	5	12	47	2	1	217

[a]Brazil (2), Mozambique (4), São Tomé and Príncipe (4), Spain (1), not stated (3).

C2V3C3 amino acid sequence analysis

Figure 2A shows the alignment of multiple amino acid sequences deduced from the C2V3C3 nucleotide sequences used for phylogenetic analysis, grouped according to each subtype consensus. Most of the strains encode V3 loops composed of 35 amino acids bounded by a conserved cysteine–cysteine disulfide bridge. In four viral strains (subtype B PT482 and PT457, subtype D PT073, and subtype G PT004) the V3 loop 3′ cysteine is missing. Seven isolates possess 34 amino acid residues (subtype A PT522; subtype B PT301 and PT695; subtype D PT073, PT418, and PT787; and subtype G PT269) in the V3 loop and one isolate has 33 residues (subtype D PT317), whereas two isolates (subtype B PT968 and subtype D PT307) have one additional residue relative to each of the subtype consensuses. At the crown of the V3 loop, 11 different sequences were found. GPGR is found predominantly (58.8%) in the 17 subtype B strains sequenced and GPGQ is the most common (75.6%) for all other subtypes. Unusual crown motifs GVPGE and PGQ were documented for subtype B and GPGT and APGT were documented for subtype G sequences. High inter- and intrasubtype diversity is observed in the V3 loop cytotoxic T lymphocyte (CTL) epitope (amino acid positions 306–325) and in the principal antibody-neutralizing determinant (PND) V3 loop crown sequence (positions 309–317). Potential N-linked glycosylation sites are relatively conserved through most of the C2 region and inside the V3 loop (positions 301–303). High variability is observed for glycosylation sites flanking the V3 loop (positions 289–291, 295–297, and 332–334). However, selective pressure seems to operate in order to maintain a minimum of glycosylation at the base of the V3 loop, because, most of the time, when one glycosylation motif is lost another emerges close to the previous one (Fig. 2A).

It is known that the emergence of HIV-1 variants with a syncytium-inducing (SI) phenotype during the natural course of infection correlates with disease progression and may be involved in AIDS pathogenesis. The macrophage-tropic, nonsyncytium-inducing (NSI) viruses are the predominant viruses found in the asymptomatic stage, whereas the T cell-tropic, SI viruses are often isolated from patients with AIDS. The majority of NSI strains tested infect CCR5 cells and possess the V3 loop amino acid sequence S/GXXXG-PGXXXXXXXE/D (X representing any amino acid).[38] SI viruses usually use CXCR4 as coreceptor, are associated with positively charged residues at positions 306 and 322, and an overall V3 loop net charge above +6.[39] Independent of the genetic subtype, most of the V3 loops sequenced have noncharged amino acids at the above-described positions and have an amino acid motif predictive of CCR5 usage. The presence of basic residues at position 306 or 322 was observed for nine of the sequences (subtype B PT893, PT709, PT508, PT947, PT170, PT079, and PT301, and subtype G PT037 and PT269). All but four sequences (PT947, PT170, PT037, and PT269) have an overall V3 loop charge equal or less than +6 (determination made with Pepstats, available online at http://bioweb.pasteur.fr/seqanal/interfaces/pepstats.html). Individuals infected with viral isolates PT508, PT037 and PT269 have already developed AIDS, whereas those infected with PT893, PT709, PT170, and PT301 are symptomatic carriers with $CD4^+$ cell counts $<200/\mu l$.

gp41 amino acid sequence analysis

Figure 2B shows alignments of the predicted gp41 N-terminal amino acid sequences, compared with consensus sequences A, B, F, and G. This region includes two immunodominant domains: cluster I (amino acids 571–614), containing the CTL epitope (positions 582–593; AVERYLKDQQLL) and the cysteine loop (positions 598–604; CSGKLIC), and cluster II (positions 637–685), containing the ectodomain region (positions 662–667; A/ELDKWA) that has been implied in viral entry.[40] Both the CTL epitope and the ectodomain sequences revealed a high degree of inter- and intrasubtype amino acid conservation. Compared with the Los Alamos amino acid G consensus, mutations T612A, M626L, I629V, I646L, and K677N (Fig. 2B) were observed for all the subtype G strains and might reflect the subcluster (supported by a bootstrap value of 891) formed by these sequences in the phylogenetic tree (Fig. 1B). The CTL epitope for the subtype A sequence presents an amino acid substitution (V583L) typical of CRF02_AG viruses.

Subtype G subclusters

From the topology of the C2V3C3-based phylogenetic tree presented in Fig. 1A it becomes clear that the great majority of the Portuguese subtype G HIV-1 sequences form two separate subclusters, tentatively named G^P and $G^{P\prime}$, supported by high bootstrap values of 925 and 980, respectively. Both groups segregate away from the subtype G reference nucleotide sequences used in our phylogenetic analysis, which include two distinct subclusters of Nigerian sequences defined by Peeters *et al.*[41] and designated GA-NG003 and G′. The subclusters were confirmed by maximum likelihood analysis (data not shown) and by multivariate principal coordinate analysis (Fig. 3). The coordinate scores obtained by PCOORD are strong, covering 26% of the total variation in the DNA sequences analyzed (14% for the first coordinate and 12% for the second), indicating a clear distinction between the two groups of subtype G HIV-1 sequences. No specific epidemiological feature, including geographic origin, age, sex or risk group, was associated with the clustering.

The DNA sequences included in the G^P subcluster, the one most distantly related to any of the other subtype G sequences, were further analyzed by SUDI, which has been developed specifically to determine whether new lineages of sequences represent new subtypes or subsubtypes. Consistent with previous reports,[3,42] analysis of the reference set yielded intrasubtype distances of 8 to 16%, intersubsubtype distances of 16 to 23%, and intersubtype distances ranging from 22 to 29%. For the newly described G^P sequences, as expected, subtype G was identified as the closest relative, whereas the distance values derived from comparisons between the new sequences and the prototypic subtype G sequences fell within the subsubtype range (15 to 22%). These data suggest that the G^P subcluster may constitute a subsubtype within the subtype G radiation.

Sequences from both groups G^P and $G^{P\prime}$ show significant amino acid differences when compared with the G consensus sequence or with each other. Statistical analysis by two-tailed Fisher exact test demonstrated that the presence of some amino acid residues at particular positions is strongly associated with the G^P subcluster. G^P strains possess four amino acid signatures at positions E268G, N295T, V333I, and K346T (indicated by

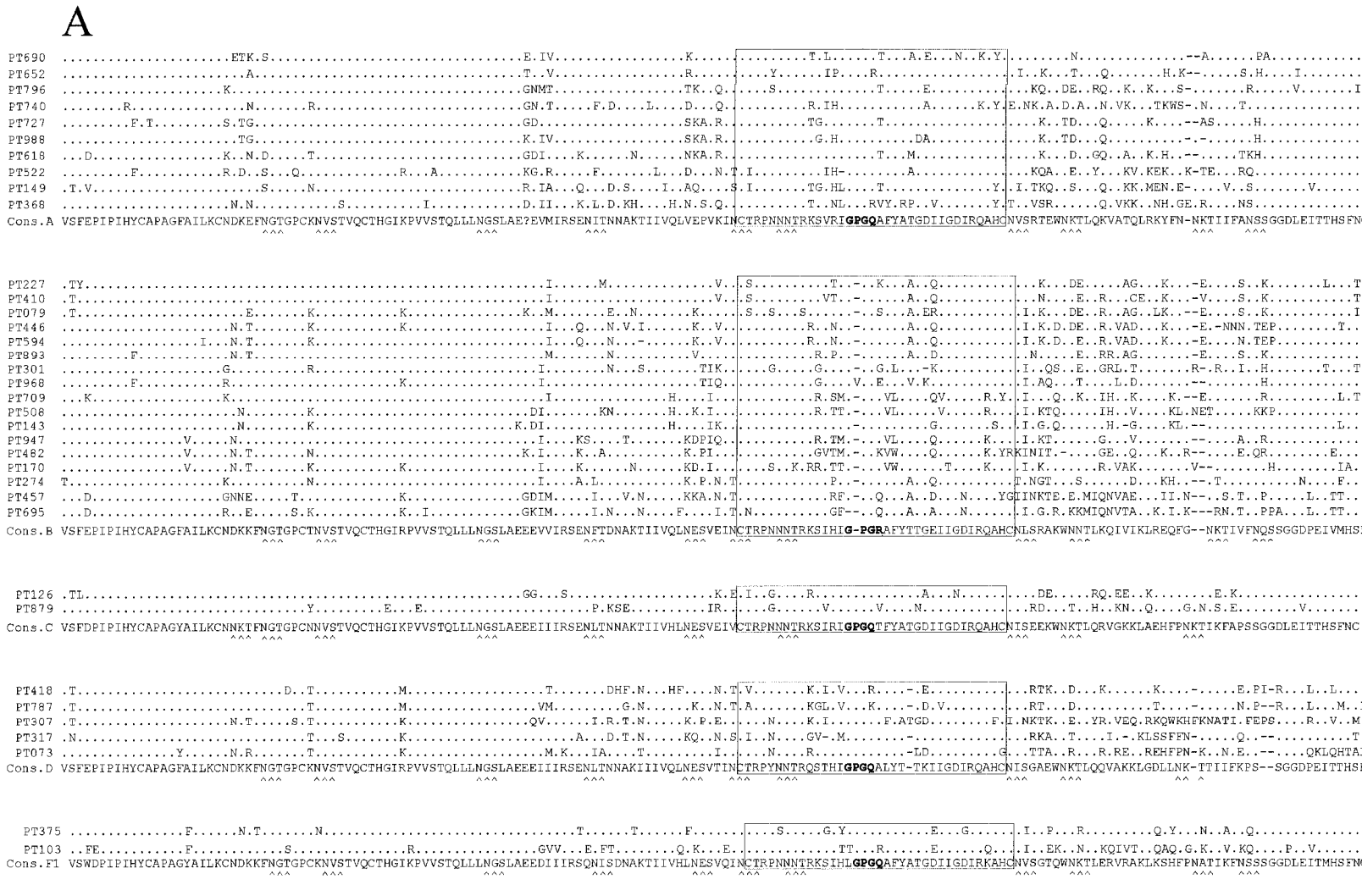


FIG. 2. Alignments of deduced protein sequences of the C2V3C3 region (**A**) and the N-terminal part of gp41 (**B**) of the sequenced Portuguese HIV-1 strains. Sequences have been grouped according to their subtype for which a consensus sequence (Cons.) has been compiled from the Los Alamos database. Dots indicate identity to the consensus, dashes correspond to gaps introduced to maintain the correct alignment, question marks indicate highly variable amino acid positions, and asterisks indicate the introduction of translation stop codons. Carets (^) show the positions of potential N-linked glycosylation sites in the consensus. The crown sequence of the V3 loop (boxed) is indicated in boldface. Within the subtype G group of sequences the subgroup formed by samples PT004 to PT694 includes those identified as G^{P}, while the PT183 to PT629 subgroup has been designated $G^{P\prime}$ (see text). The vertical arrows indicate the positions that correspond to amino acid signatures in the G^{P} subgroup. The first and last amino acids shown in (**A**) and (**B**), respectively, correspond to positions 208–378 and 563–685 in the subtype B HXB2 Env sequence. The amino acid sequences in the gp41 region (**B**), corresponding to the CTL epitope (positions 582 to 593) and the ectodomain region (positions 662 to 667), are underlined. Sequence PT068 was excluded from the gp41 amino acid alignment due to an occurrence of frameshifting.

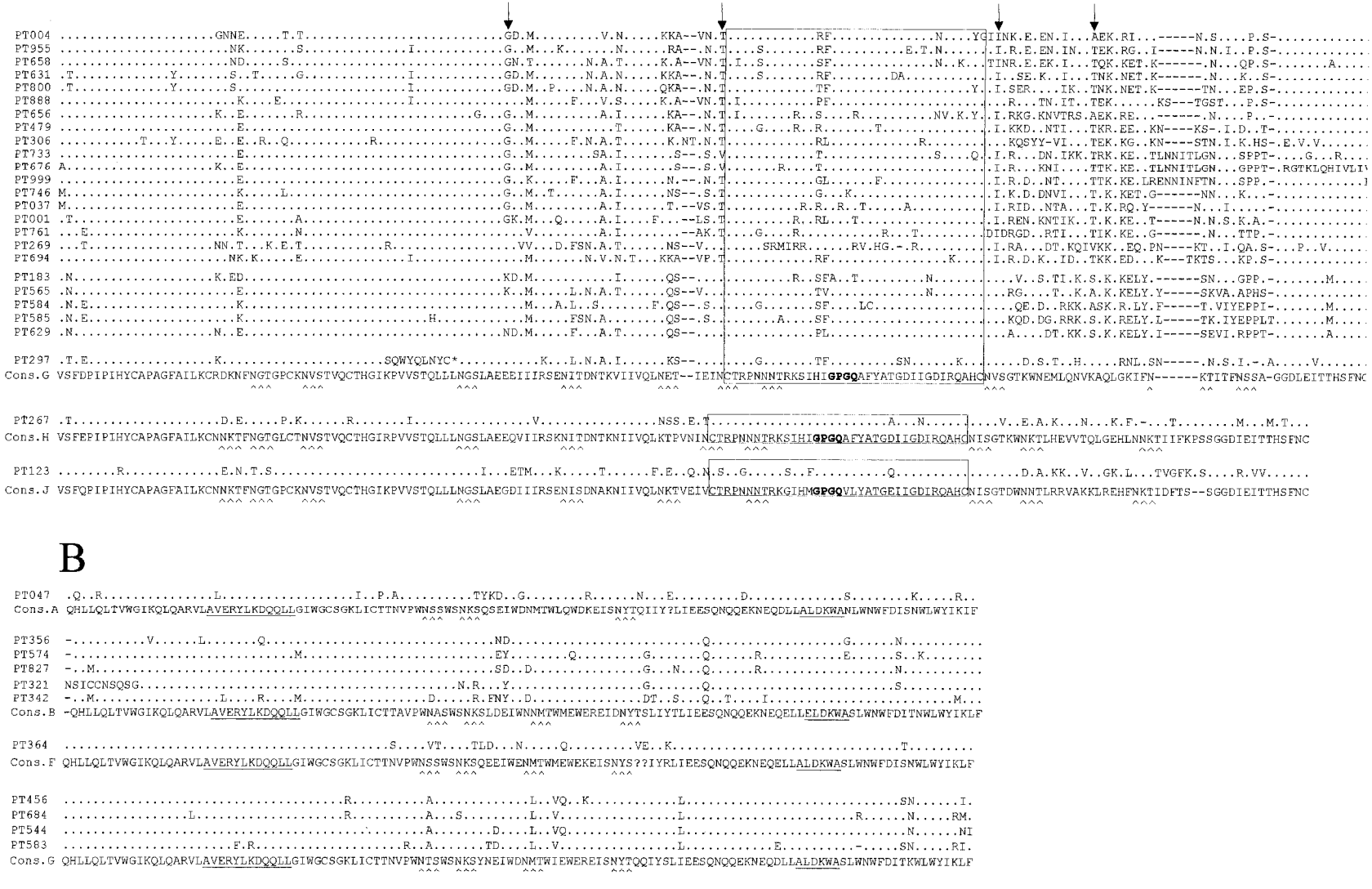


FIG. 2. *Continued.*

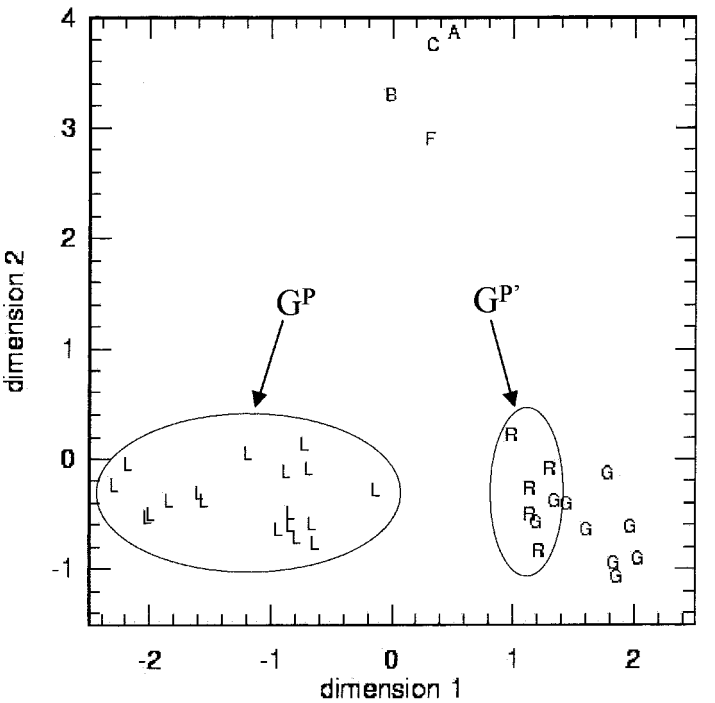


FIG. 3. Results of multivariate principal coordinate analysis (PCOORD) of the groups G^P and $G^{P\prime}$. The axes are the two dimensions that were first extracted; together they cover 26% of the total differences between the two groups. The first 10 axes cover 66% of variation. A (Q23CXCCG), B (HXB2), C (92BR025), and F (93BR020.1) represent corresponding subtype reference sequences. L and R indicate HIV-1 nucleotide sequences that have been included in the groups G^P and $G^{P\prime}$, respectively. G corresponds to HIV-1 subtype G reference sequences (RU131, NG0832, VI525, 96NG-LCSW218, 96NG-MACSW085, 96NG-IKCSW22, KOP9, NG003, and LBV21-7).

vertical arrows in Fig. 2A). Mutation N295T suppresses the potential N-linked glycosylation site (NCT) at the 5′ border of the V3 loop. Two mutations, S209N and N344K, and motifs NQS and LYN at amino acid positions 289–291 and 352–354 are characteristic of $G^{P\prime}$ sequences.

Nine isolates from the main cluster G^P were PCR amplified and sequenced in the p24/p7-coding region. Phylogenetic analysis (data not shown) confirmed the subclustering observed for the C2V3C3 sequences (bootstrap value of 994). Gag putative amino acid alignment did not show significant differences when compared with the subtype G amino acid consensus. However, at the nucleotide level, on comparison with the subtype G nucleotide consensus, eight absolutely conserved silent mutations were found, and four others were present in 88.9% of the sequences (data not shown).

DISCUSSION

In this study, we evaluated the diversity of HIV-1 envelope subtypes circulating in Portugal. The subtype screening of 241 HIV-1 strains was successfully accomplished by HMA for 143 samples and by phylogenetic analysis for 74 samples. Twenty-four specimens (10%) were systematically nonamplifiable by PCR. HIV-1 subtype B is the most prevalent but seven non-B subtypes (A, C, D, F, G, H, and J) were found in half of the cases studied.

The high level of genetic diversity detected among HIV-1 strains circulating in Portugal may account, at least in part, for the high numbers of HIV-1 strains with indeterminate HMA profiles and of samples not amplified by the PCR primers included in the standard gp120-based HMA kit. One-third of the latter samples were amplified with more generic gp41 primers. Difficulties of PCR amplification with gp120 primers included in the HMA kit and ambiguous migration profiles for subtype A (IbNG-like) and G viruses were previously observed in molecular epidemiology studies performed in Nigeria[41,43] and Democratic Republic of Congo.[44]

Our results, showing subtype B predominance in Portugal and cocirculation with non-B strains is consistent with the HIV-1 subtype distributions described for other European countries.[5–30] However, the Portuguese population studied shows a much higher prevalence of non-B subtypes. The majority of these viruses were found to infect African immigrants, but the prevalence among Portuguese-born individuals was 33.1%, which is remarkably high. This value is probably not representative of the Portuguese white population as a whole and might be biased by local proximity and social interactions with African communities, but still we should be aware of the easy introduction and spread of non-B strains and their practical implications in dealing with the epidemic.

This is the first study reporting HIV-1 subtypes from a significant number of samples from Angola ($n = 25$) and Cape Verde Islands ($n = 34$). Our finding of subtypes A, B, C, D, F,

G, H, and J in Angolan individuals is consistent with those reported for the northern neighboring countries of Congo[45] and Democratic Republic of Congo.[44,46] These data suggest that the epidemic in Angola follows a pattern closer to that described for West/Central Africa, rather than that characteristic of southern Africa, where HIV-1 subtype C dominates. Until now, the sporadic cases of genetic characterization of HIV-1 strains associated with Angola included subtypes F, G, and H.[11,47] Individuals originating from Cape Verde Islands are infected mainly with subtypes G, B, and A. It is noteworthy that about half the subtype F strains screened were found among this group. As previously reported,[48] HIV-1 subtype A is prevalent in Guinea-Bissau natives.

Most of the subtype A samples analyzed seem to be closed related to CRF02_AG *env* sequences, which are prevalent in West Africa.[49] Three subtype B samples (PT482, PT947, and PT170) formed monophyletic clusters, supported by bootstrap values of 922 (data not shown), with V3 sequences from BG recombinant viruses identified in Galicia[12] and with the Portuguese BG recombinant isolate HC10.[50] Taken together, these two findings indicate that a still undetermined proportion of viruses circulating in Portugal might be recombinant forms, which further increases the genetic diversity of Portuguese strains.

In our analysis, we have found a broad spectrum of sequences at the tip of the V3 loop. Eleven different motifs were represented in a total of 62 sequences analyzed, including some unusual ones (GVPGE, PGQ, GPGT, and APGT). The V3 loop is an important functional and biological domain for virus neutralization, cell tropism, and syncytium-inducing capabilities. The high variability observed should, therefore, be taken into account in interventions relying on V3 peptide epitopes.

Analysis of the V3 loop characteristics correlated with syncytium formation and coreceptor usage demonstrated that most sequences are typical of NSI/R5 viruses, whereas nine of them might belong to the SI/X4 type. Seven subjects infected with these viruses developed AIDS or are symptomatic carriers with low $CD4^+$ cell counts, demonstrating a good correlation between the predicted viral properties and disease progression.

The finding that Portuguese sequences from the most prevalent subtypes (A, B, and G) are not clustered together suggests multiple introductions of those viruses. Interestingly, two clusters of HIV-1 subtype G isolates were seen; one, $G^{P\prime}$, is more closely related to the prototypic G strains than the other, G^P, which shows distant genetic relationships to the reference sequences in the tree and could represent a new G subsubtype. The confirmation of these results will wait until the full genome sequence becomes available.

In conclusion, at least eight subtypes (A, B, C, D, F, G, H, and J) and, most probably, two CRFs define the HIV-1 infection in Portugal. Predominance of subtype B was observed but non-B strains, as a whole, were identified in half of the samples analyzed. Our results emphasize the change in the current global distribution of HIV-1 variants, mainly through immigration. The genetic diversity observed could be responsible for the difficulties with HMA subtyping. Introduction of new primers/reference strains in the gp120-HMA kit and/or the use of the gp41-HMA approach[43] might be useful to minimize these problems. Circulation of non-B strains could have significant implications for the Portuguese population, both for the efficacy of antiretroviral therapy and viral load assays. In the absence of well-established correlates of immunity or protection, the issue of viral subtype diversity will continue to be an important consideration in the design of appropriate vaccines to elicit broad protective immunity against all known HIV-1 subtypes. In light of the high diversity of HIV-1 *env* subtypes in Portugal, our data reinforce the need for continuous epidemiological surveillance and further analysis of other genome regions.

ACKNOWLEDGMENTS

We thank the staff of the Serviço de Patologia Clínica of the Hospital Fernando da Fonseca, especially Mrs. Teresa Gil, for their dedicated cooperation. Part of this study was financially supported by the Japan Health Sciences Foundation (Tokyo, Japan).

SEQUENCE DATA

The nucleotide sequences reported in this study have been submitted to the EMBL sequence database under the following accession numbers: AJ296216, AJ296217, AJ296219–AJ296267, and AJ318386–AJ318396 (C2V3C3 sequences); AJ306161, AJ306162, and AJ318397–AJ318406 (gp41 sequences); AJ306149–AJ306160 (*gag* sequences).

REFERENCES

1. UNAIDS/WHO: *UNAIDS/WHO Epidemiological Fact Sheet on HIV/AIDS and Sexually Transmitted Infections, Portugal, 2000 Update*.http://www.unaids.org/hivaidsinfo/statistics/june00/fact_sheets/pdfs/ portugal.pdf.
2. Comissão Nacional de Luta Contra a SIDA: *SIDA, a Situação em Portugal a 30 de Junho de 2000*. Doc. 121. Centro de Vigilância Epidemiológica das Doenças Transmissíveis, Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisbon, Portugal, 2000.
3. Robertson DL, Anderson JP, Bradac JA, *et al.*: HIV-1 nomenclature proposal: A reference guide to HIV-1 classification. In: *Human Retroviruses and AIDS* (Kuiken C, Foley B, Hahn B, *et al.*, eds.) Los Alamos National Laboratory, Los Alamos, New Mexico, 1999, pp. 492–505.
4. McCutchan FE: Understanding the genetic diversity of HIV-1. AIDS 2000;14:S31–S44.
5. Lasky M, Perret JL, Peeters M, *et al.*: Presence of multiple non-B subtypes and divergent subtype B strains of HIV-1 in individuals infected after overseas deployment. AIDS 1997;11:43–51.
6. Couturier E, Damond F, Roques P, *et al.*: HIV-1 diversity in France, 1996–1998. AIDS 2000;14:289–296.
7. Alaeus A, Leitner T, Lidman K, and Albert J: Most HIV-1 genetic subtypes have entered Sweden. AIDS 1997;11:199–202.
8. Iversen AK, Learn GH, Fugger L, Gerstoft J, Mullins JI, and Skinhoj P: Presence of multiple HIV subtypes and a high frequency of subtype chimeric viruses in heterosexually infected women. J Acquir Immune Defic Syndr 1999;22:325–332.
9. Böni J, Pyra H, Gebhardt M, *et al.*: High frequency of non-B subtypes in newly diagnosed HIV-1 infections in Switzerland. J Acquir Immune Defic Syndr 1999;22:174–179.
10. Parry JV, Murphy G, Barlow KL, *et al.*: National surveillance of

HIV-1 subtypes for England and Wales. Design, methods and initial findings. J Acquir Immune Defic Syndr 2001;26:381–388.
11. Holguín A, Rodés B, and Soriano V: Protease gene analysis of HIV type 1 non-B subtypes in Spain. AIDS Res Hum Retroviruses 2000;16:1395–1403.
12. Thomson MM, Delgado E, Manjón N, *et al.*: HIV-1 genetic diversity in Galicia, Spain: BG intersubtype recombinant viruses circulating among injecting drug users. AIDS 2001;15:509–516.
13. Nasioulas G, Paraskevis D, Paparizos V, Lazanas M, Karafoulidou A, and Hatzakis A for the Multicentre Study on HIV-1 Heterogeneity: Genotypic characterization of human immunodeficiency virus type 1 in Greece. AIDS Res Hum Retroviruses 1998;14: 685–690.
14. Kostrikis LG, Bagdades E, Cao Y, Zhang L, Dimitriou D, and Ho DD: Genetic analysis of human immunodeficiency type 1 strains from patients in Cyprus: Identification of a new subtype designated subtype I. J Virol 1995;69:6122-6130.
15. Heyndrickx L, Janssens W, Coppens S, *et al.*: HIV type 1 C2V3 *env* diversity among Belgian individuals. AIDS Res Hum Retroviruses 1998;14:1291–1296.
16. Op de Coul E, van der Schoot A, Goudsmit J, *et al.*: Independent introduction of transmissible F/D recombinant HIV-1 from Africa into Belgium and The Netherlands. Virology 2000;270:267–277.
17. Dietrich U, Ruppach H, Gehring S, *et al.*: Large proportion of non-B HIV-1 subtypes and presence of zidovudine resistance mutations among German seroconverters. AIDS 1997;11:1532–1533.
18. Romano L, Venturi G, Ferruzzi R, *et al.*: Detection of genotypically drug-resistant HIV-1 variants and non-B subtypes in recently infected antiretroviral-naive adults in Italy. AIDS 2000;14: 2204–2206.
19. Puchhammer-Stöckl E, Kunz C, Faatz E, Kasper P, and Heinz FX: Introduction of HIV-1 subtypes C, E and A into Austria. Clin Diagn Virol 1998;9:25–28.
20. Jonassen TO, Grinde B, Åsjö B, Hasle G, and Hungnes O: Intersubtype recombinant HIV type 1 involving HIV-MAL-like and subtype H-like sequence in four Norwegian cases. AIDS Res Hum Retroviruses 2000;16:49–58.
21. Bobkov A, Cheingsong-Popov R, Garaev M, *et al.*: Identification of an *env* G sutype and heterogeneity of HIV-1 strains in the Russian Federation and Belarus. AIDS 1994;8:1649–1655.
22. Bobkov A, Cheingsong-Popov R, Selimova L, *et al.*: Genetic heterogeneity of HIV type 1 in Russia: Identification of H variants and relationship with epidemiological data. AIDS Res Hum Retroviruses 1996;12:1687–1690.
23. Bobkov A, Cheingsong-Popov R, Selimova L, *et al.*: HIV type 1 subtype E in Russia. AIDS Res Hum Retroviruses 1997;13: 725–727.
24. Lukashov VV, Huisman R, Rakhmanova AG, *et al.*: Circulation of subtype A and gagA/envB recombinant HIV type 1 strains among injecting drug users in St. Petersburg, Russia, correlates with geographical origin of infections. AIDS Res Hum Retroviruses 1999;15:1577–1583.
25. Liitsola K, Laukkanen T, Denisova A, *et al.*: Genetic characterization of HIV-1 strains in the Baltic countries and Russia. Scand J Infect Dis 1996;28:537–541.
26. Apetrei C, Necula A, Holm-Hansen C, *et al.*: HIV-1 diversity in Romania. AIDS 1998;12:1079–1085.
27. Mezei M, Balog K, Takács M, *et al.*: Genetic subtypes of HIV type 1 in Hungary. AIDS Res Hum Retroviruses 2000;16:513–516.
28. Hampl H, Sawitzky D, Stoffler-Meilicke M, *et al.*: First case of HIV-1 subtype O infection in Germany. Infection 1995;23: 369–370.
29. Peeters M, Lobe V, Nkengasong J, *et al.*: HIV-1 group O infection in Belgium. Acta Clin Belg 1995;50:171–173.
30. Garcia-Lerma G, Gutierrez M, Mas A, Bravo R, Aguilera O, and Soriano V: Report of the first 2 cases of HIV-1 group O infection in Spain. Med Clin (Barc) 1996;107:418–421.
31. Heredia A, Vallejo A, Soriano V, *et al.*: Phylogenetic analysis of HIV type 2 strains from Portugal. AIDS Res Hum Retroviruses 1998;14:471–473.
32. Yang C, Dash BC, Simon F, *et al.*: Detection of diverse variants of human immunodeficiency virus-1 groups M, N, and O and simian immunodeficiency viruses from chimpanzees by using generic *pol* and *env* primer pairs. J Infect Dis 2000;181:1791–1795.
33. Fransen K, Zhong P, De Beenhouwer H, *et al.*: Design and evaluation of new, highly sensitive and specific primers for polymerase chain reaction detection of HIV-1 infected primary lymphocytes. Mol Cell Probes 1994;8:317–322.
34. Delwart EL, Shpaer EG, Louwagie J, *et al.*: Genetic relationships determined by a DNA heteroduplex mobility assay: Analysis of HIV-1 *env* genes. Science 1993;262:1257–1261.
35. Heyndrickx L, Janssens W, Zekeng L, *et al.*: Simplified strategy for detection of recombinant human immunodeficiency virus type 1 group M isolates by *gag/env* heteroduplex mobility assay. J Virol 2000;74:363–370.
36. Hasegawa M, Kishino H, and Yano T: Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 1985;22:160–174.
37. Page RD: TreeView: An application to display phylogenetic trees on personal computers. Comput Appl Biosci 1996;12:357–358.
38. Xiao L, Owen SM, Goldman I, *et al.*: CCR5 coreceptor usage of non-syncytium-inducing primary HIV-1 is independent of phylogenetically distinct global isolates: Delineation of consensus motif in the V3 domain that predicts CCR5–usage. Virology 1998; 240:83–92.
39. Kim FM, Kolson DL, Balliet JW, Srinivasan A, and Collman RG: V3–independent determinants of macrophage tropism in a primary human immunodeficiency virus type 1 isolate. J Virol 1995; 69:1755–1761.
40. Dorn J, Masciotra S, Yang C, *et al.*: Analysis of genetic variability within the immunodominant epitopes of envelope gp41 from human immunodeficiency virus type 1 (HIV-1) group M and its impact on HIV-1 antibody detection. J Clin Microbiol 2000;38: 773–780.
41. Peeters M, Esu-Williams E, Vergne L, *et al.*: Predominance of subtype A and G HIV type 1 in Nigeria, with geographical differences in their distribution. AIDS Res Hum Retroviruses 2000;16: 315–325.
42. Gao F, Vidal N, Li Y, *et al.*: Evidence of two distinct subsubtypes within the HIV-1 subtype A radiation. AIDS Res Hum Retroviruses 2001;17:675–688.
43. Agwale SM, Robbins KE, Odama L, *et al.*: Development of an *env* gp41-based heteroduplex mobility assay for rapid human immunodeficiency virus type 1 subtyping. J Clin Microbiol 2001;39: 2110–2114.
44. Vidal N, Peeters M, Mulanga-Kabeya C, *et al.*: Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. J Virol 2000;74:10498–10507.
45. Bikandou B, Takehisa J, Mboudjeka I, *et al.*: Genetic subtypes of HIV type 1 in Republic of Congo. AIDS Res Hum Retroviruses 2000;16:613–619.
46. Yang C, Dash B, Hanna SL, *et al.*: Predominance of HIV type 1 subtype G among commercial sex workers from Kinshasa, Democratic Republic of Congo. AIDS Res Hum Retroviruses 2001; 17:361–365.
47. Holguín A, Rodés B, Dietrich U, and Soriano V: Human immu-

nodeficiency viruses type 1 subtypes circulating in Spain. J Med Virol 1999;59:189–193.
48. Esteves A, Parreira R, Piedade J, Venenno T, and Canas-Ferreira WF: Genetic characterization of HIV type 1 and type 2 from Bissau, Guinea-Bissau (West Africa). Virus Res 2000;68:51–61.
49. Montavon C, Toure-Kane C, Liegeois F, *et al.*: Most *env* and *gag* subtype A HIV-1 viruses circulating in West and West Central Africa are similar to the prototype AG recombinant virus IBNG. J Acquir Immune Defic Syndr 2000;23:363–374.
50. Villahermosa ML, Thomson M, Vázquez de Parga E, *et al.*: Improved conditions for extraction and amplification of human immunodeficiency virus type 1 RNA from plasma samples with low viral load. J Hum Virol 2000;3:27–34.

Address reprint requests to:
Aida Esteves
Instituto de Higiene e Medicina Tropical (UNL)
Unidade de Virologia
Rua da Junqueira, 96
P-1349–008 Lisbon, Portugal

E-mail: aidaesteves@ihmt.unl.pt