

## Sequence Note

# Spreading of HIV-1 Subtype G and *envB/gagG* Recombinant Strains among Injecting Drug Users in Lisbon, Portugal

AIDA ESTEVES,<sup>1</sup> RICARDO PARREIRA,<sup>1</sup> JOÃO PIEDADE,<sup>1</sup> TERESA VENENNO,<sup>1</sup>  
MARGARIDA FRANCO,<sup>2</sup> JOSÉ GERMANO DE SOUSA,<sup>2</sup> LUIS PATRÍCIO,<sup>3</sup> PAULA BRUM,<sup>3</sup>  
ANTÓNIO COSTA,<sup>3</sup> and WANDA F. CANAS-FERREIRA<sup>1</sup>

### ABSTRACT

We have evaluated the genetic diversity of HIV-1 strains infecting injecting drug users (IDUs) in Lisbon, Portugal. Heteroduplex mobility assay and/or phylogenetic analysis revealed that *env* (C2V3C3 or gp41) subtype B is present in 63.7% of the 135 viral samples studied, followed by subtypes G (23.7%), A (6.7%), F (5.2%), and D (0.7%). Similar analysis of *gag* (p24/p7) performed on 91 of the specimens demonstrated that 49.5% of the infections were caused by subtype G viruses; other *gag* subtypes identified were B (39.5%), F (3.3%), A and D (1.1% each), and the recombinant circulating form CRF02\_AG (5.5%). Discordant *env/gag* subtypes were detected in 34.1% of the strains and may reflect the presence of dual infections and/or recombinant viruses. The presumptive B/G recombinant form was highly predominant (21 of 31). The genetic pattern of HIV-1 subtype B and G strains is suggestive of multiple introductions and recombination episodes and of a longstanding presence of both subtypes in the country. C2V3C3 amino acid sequences from IDU-derived subtype G viruses presented highly significant signatures, which distinguish the variants from this transmission group. The unusually high prevalence of subtype G sequences (34.1%), independent of the geographic origin of the infected individuals, makes this IDU HIV-1 epidemic unique.

INJECTING DRUG USERS (IDUs) form one of the major risk groups for HIV-1 infection in Portugal. In the period of 1997–2001, 62% of the reported AIDS cases were related to this transmission class<sup>1</sup> but, following the general trends in most Western European countries, the incidence of HIV-1 infection among Portuguese IDUs has started to decrease.

Phylogenetic analyses of HIV-1 isolates distributed worldwide showed, overall, 23 distinct HIV-1 subtypes (A–D, F–H, J, and K) and circulating recombinant forms (CRF01–14) within group M (HIV Sequence Database, <http://hiv-web.lanl.gov/>). These HIV-1 subtypes are geographically distributed nonrandomly, with subtype B being predominant in Europe and North America. We have reported multiple *env* HIV-1 subtypes circulating in Portugal, with a high prevalence of non-B subtypes, both in African and Portuguese individuals.<sup>2</sup> The aim of the

present study was to investigate the genetic diversity of incident HIV-1 strains among IDUs living in Lisbon.

From May 1999 through December 2000, after obtaining informed consent, blood samples were collected from 81 HIV-1 seropositive individuals attending three different drug addiction recovery centers and enrolled in a methadone therapy program. Seventy additional seropositive IDUs had been part of a previous broader surveillance study.<sup>2</sup> Among these 151 IDUs, the mean age was 31 years, ranging from 18 to 49 years; the great majority (88.1%) were of Portuguese nationality, and 29.1% were women.

Proviral DNA from uncultured peripheral blood mononuclear cells was extracted and we proceeded to the analysis of the *env* C2V3C3 (nucleotides 6982–7246, based on HXB2) or the gp41 (nucleotides 7878–8282 of HXB2) regions by het-

<sup>1</sup>Unidade de Virologia/Unidade de Parasitologia e Microbiologia Médicas, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, P-1349-008 Lisbon, Portugal.

<sup>2</sup>Laboratório de Patologia Clínica, Hospital Fernando da Fonseca, P-2700-276 Amadora, Portugal.

<sup>3</sup>Centro de Atendimento a Toxicodependentes das Taipas, P-1250 Lisbon, Portugal.

eroduplex mobility assay (HMA) and/or DNA sequencing followed by phylogenetic analysis. The same cellular lysates were used to amplify by polymerase chain reaction (PCR) a product of 464 bp in the p24-p7 region (nucleotides 1577–2040 of HXB2) for use in *gag* HMA.<sup>3</sup> The conditions for DNA extraction, nested PCR amplification, HMA, and DNA sequencing were as previously described.<sup>2,3</sup> DNA and predicted protein sequences were aligned using CLUSTAL W software (available at <http://www.ebi.ac.uk/clustalw/>). Pairwise genetic distances were calculated with the Kimura two-parameter algorithm and tree topologies were inferred by the neighbor-joining method (using DNADIST) as implemented in the PHYLIP version 3.4 software package (available at <http://bioweb.pasteur.fr>). Phylogenetic trees were drawn with TREEVIEW<sup>4</sup> and their robustness was evaluated by bootstrap analysis of 1000 resamplings, using the SEQBOOT and CONSENSUS programs.

One hundred and thirty-five samples could be successfully amplified with C2V3C3-specific ( $n = 124$ ) or gp41-specific ( $n = 11$ ) primers, whereas 16 samples, despite repeated attempts, were not amplified in either region. HMA subtyping performed with the C2V3C3 amplification products revealed that 77 individuals were infected with HIV-1 subtype B, 30 with G, 7 with A, and 5 with F viral strains. On the basis of phylogenetic analysis, five samples with undetermined HMA profiles were subtyped as CRF02\_AG ( $n = 2$ ), F ( $n = 2$ ), and D ( $n = 1$ ) (Fig. 1). Nine additional individuals were found to be infected with subtype B viruses and two with G strains, as indicated by phylogenetic analysis of gp41 sequences (data not shown; DDBJ/EMBL/GenBank database accession numbers AJ318398–AJ318400, AJ306161, AJ306162, AJ318404, and AJ429038–AJ429042). On the whole, the majority (63.7%) of the viral samples analyzed belong to subtype B and the remaining non-B specimens include viruses from *env* subtypes G (23.7%), A (6.7%), F (5.2%), and D (0.7%).

From the 135 HIV-1 samples previously subtyped for *env*, 91 (67.4%) were amplified for *gag* and the respective DNA products were used for HMA. Subtypes A, B, D, F, G, and CRF02\_AG were identified unequivocally in 85 samples, using this approach. Six of the amplified fragments that could not be subtyped by *gag* HMA were sequenced and the viral subtype was inferred by phylogenetic analysis (data not shown; DDBJ/EMBL/GenBank database accession numbers AJ504457–AJ504462). Altogether, viral samples were classified for *gag* as subtypes G (49.5%), B (39.5%), CRF02\_AG (5.5%), F (3.3%), A (1.1%), and D (1.1%). There was accordance between the *env* and *gag* subtypes in 60 samples (35 B/B, 22 G/G, 2 F/F, and 1 D/D). About one-third of the samples (31 of 91) showed discordant subtypes, suggesting the presence of recombinant viruses and/or dual infections.

Table 1 describes the potential recombinant forms found and their relative frequency. B/G (*env/gag*) strains are the most prevalent (67.7%), and, at least to our knowledge, three putative new genetic forms were found with subtypes F/G, B/A (IbNG), and G/B. The *env* subtype A viruses also analyzed in *gag* (four of nine) were all identified as CRF02\_A/G. All the double-subtyped *env* G strains, with one exception, were also classified as subtype G in *gag*. The high prevalence (24.2%) of potential nonrecombinant G viruses described in this work, as well as the *env* HIV-1 subtyping data for Portugal previously reported by us,<sup>2</sup> is suggestive of a Portuguese origin for most G recombinants circulating in the Iberian Peninsula.<sup>5</sup> Altogether, non-B viral sequences were identified in 46.0% of the genes analyzed, 74.0% of which belong to subtype G. Among the 121 specimens amplified from native Portuguese IDUs, 54.5% were of non-B subtype, either in *gag* or in *env*. Subtype distribution, B versus non-B and G versus non-G, was found to be independent of gender ( $\chi^2$  test with the Yates correction,  $p > 0.01$ ) and geographic origin (two-tailed Fisher exact test,  $p > 0.01$ ).

A subset of the Portuguese IDU samples, including all subtype G and 28 randomly selected subtype B viruses, was chosen for further C2V3C3 DNA sequencing and phylogenetic analysis. Phylogenetic analysis (Fig. 1) demonstrated a 100% correlation with the previous HMA subtyping for subtype B viruses and confirmed subtyping of G strains with low-mobility heteroduplexes on HMA gels. The B subtype samples of Portuguese IDUs clustered with corresponding B reference sequences with a bootstrap value of 991. Three-quarters of the B sequences formed two major separate subclusters (I and II in Fig. 1), whereas the remaining sequences were distributed among the branches of the tree corresponding to the references. Each cluster, although supported by a relatively low bootstrap value (792 and 638 for clusters I and II, respectively), could be further substantiated by the presence of a distinct putative amino acid signature pattern (Fig. 2). Within the major B subclusters three minor groups (samples PT5866i/PT5549i/PT709, PT410/PT227/PT5192i, and PT446/PT594/PT6068i) include highly related sequences. However, besides sample pair PT5549i/PT5866i, provided by individuals who live in close proximity, making needle sharing plausible, no epidemiological link could be established between these individuals. Samples PT4789i, PT2272i, and PT170 form a monophyletic group (bootstrap value of 963) with Spanish HIV-1 isolates (subcluster III in Fig. 1) described as B/G recombinants (CRF14\_BG).<sup>6</sup> Their corresponding amino acid sequences share a four-amino acid signature pattern (group III in Fig. 2)

**FIG. 1.** Phylogenetic analysis of HIV-1 strains from Lisbon IDUs. The phylogenetic tree (neighbor joining) was constructed from alignments in which representative HIV-1 reference strains from groups M (subtypes A to K, as indicated), O (ANT70), and N (YBF30) were included and SIVcpzANT was regarded as the outgroup sequence. Genetic distance matrixes were calculated by the Kimura two-parameter method, based on 502 unambiguously aligned nucleotides covering the *env* C2V3C3 genome region. The DNA sequences described in this report are indicated in boldface, whereas those highlighted in gray represent G(*gag*)/B(*env*) recombinant viruses. In the G clade, the G<sup>P</sup> subgroup corresponds to the one previously described.<sup>2</sup> I, II, and III stand for different groups of clade B. Branch nodes supported by bootstrap values above 750 (of 1000 resamplings) are indicated. The scale bar represents 10% genetic diversity. The sequences were submitted to the DDBJ/EMBL/GenBank database under the following accession numbers: AJ296217, AJ296219–AJ296221, AJ296224, AJ296226, AJ296227, AJ296233, AJ296236, AJ296239, AJ296-245, AJ296248, AJ296251, AJ296255, AJ296262, AJ296264, AJ318390, AJ318396, AJ428995–AJ429007, AJ429009–AJ429-034, AJ429036, AJ429037, and AJ504453–AJ504456.

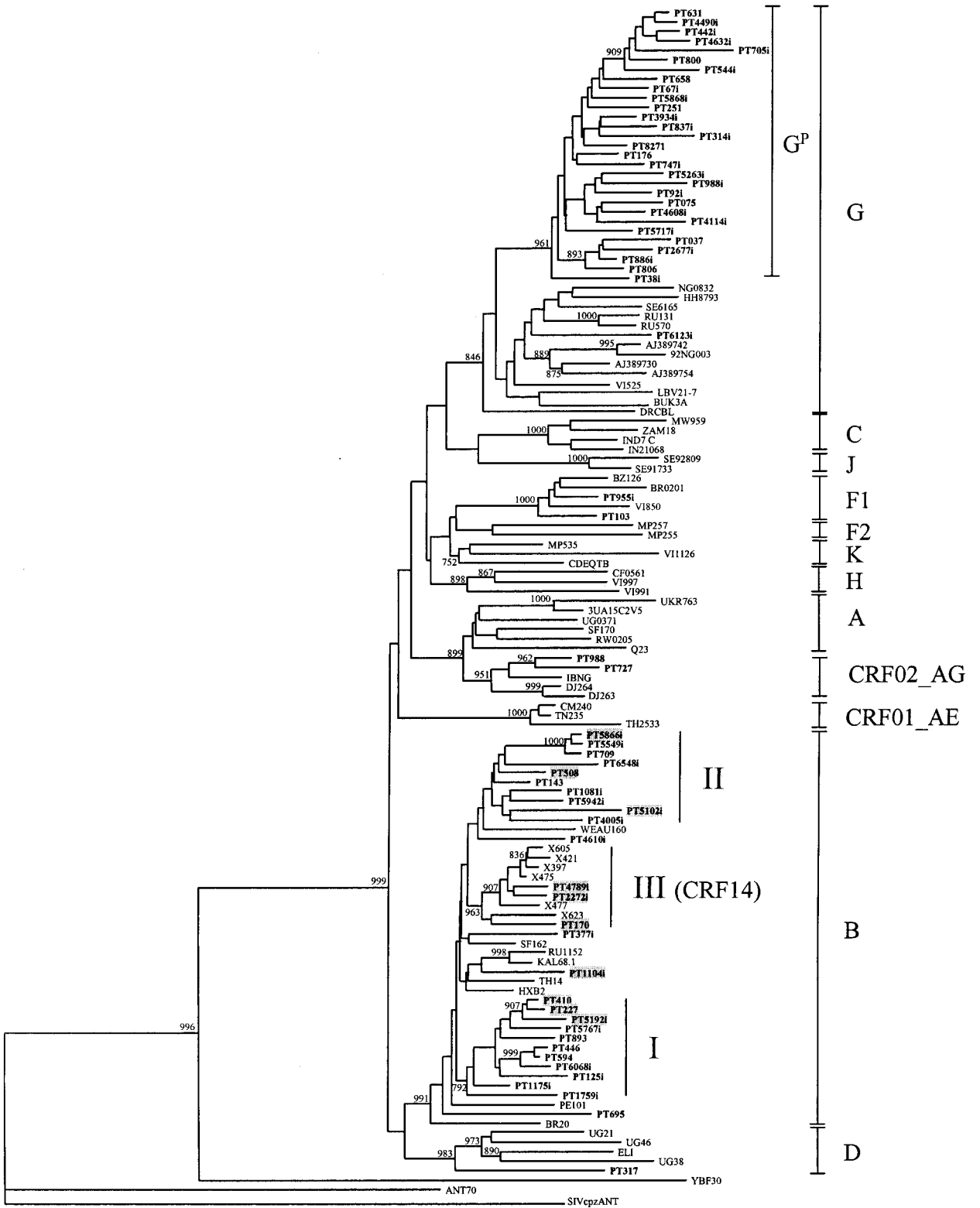


TABLE 1. POTENTIAL HIV-1 INTERSUBTYPE RECOMBINANTS IN IDUs FROM LISBON, PORTUGAL

Subtype		Number of samples
env	gag	
A	A (IbNG)	4
B	A	1
	A (IbNG)	1
	F	1
	G	21
F	G	2
G	B	1
Total:		31

not found in the other B strains analyzed. The recombinant character of these viruses was confirmed by *gag* HMA. Other B/G recombinants (shaded in gray in Fig. 1) are distributed along the B cluster, suggesting the occurrence of multiple recombination episodes. This hypothesis is supported by the high value of mean nucleotide distance ( $12.6 \pm 3.2\%$ , ranging from 3.1 to 17.3%) between the *env* sequences of B/G recombinants. Sample PT709 is *env* subtype B and *gag* subtype A. An A/B recombinant form (CRF03\_AB) circulating among IDUs in Russia was previously described.<sup>7,8</sup> However, the presently described A/B recombinant might be a different one because its *env* sequence clusters further apart from *env* reference sequences RU1152 and KAL68.1. As a whole, the branching pattern observed for subtype B sequences is suggestive of multiple B virus introductions into the Portuguese IDU population studied.

The interindividual heterogeneity of *env* subtype B specimens was on average  $12.4 \pm 3.2\%$  (1.2 to 19.9%) at the nucleotide level. Three synonymous nucleotide substitutions, **GGC** in the second glycine codon at the tip of the V3 loop, **AGG** and **TCC** at HIV-1<sub>LAI</sub> *env* positions 834 and 837, have been associated with IDU subtype B viruses from Northern Europe,<sup>9,10</sup> the United States,<sup>9</sup> Greece,<sup>11</sup> and Switzerland.<sup>12</sup> Only 2 and 1 of the 28 B strains analyzed were **GGC** and **TCC** viruses, respectively. The silent substitution **AGG** was found in 12 of the 28 IDU B sequences analyzed (10 of them included in group I). However, three of seven B sequences from individuals with a reported heterosexual route of HIV-1 transmission also presented this mutation<sup>2</sup> (data not shown). These results indicate that the HIV-1 B epidemic among Portuguese IDUs was established from diverse sources, most probably as in other risk groups, some of them differing from the founder strains proposed for other Western European IDUs.

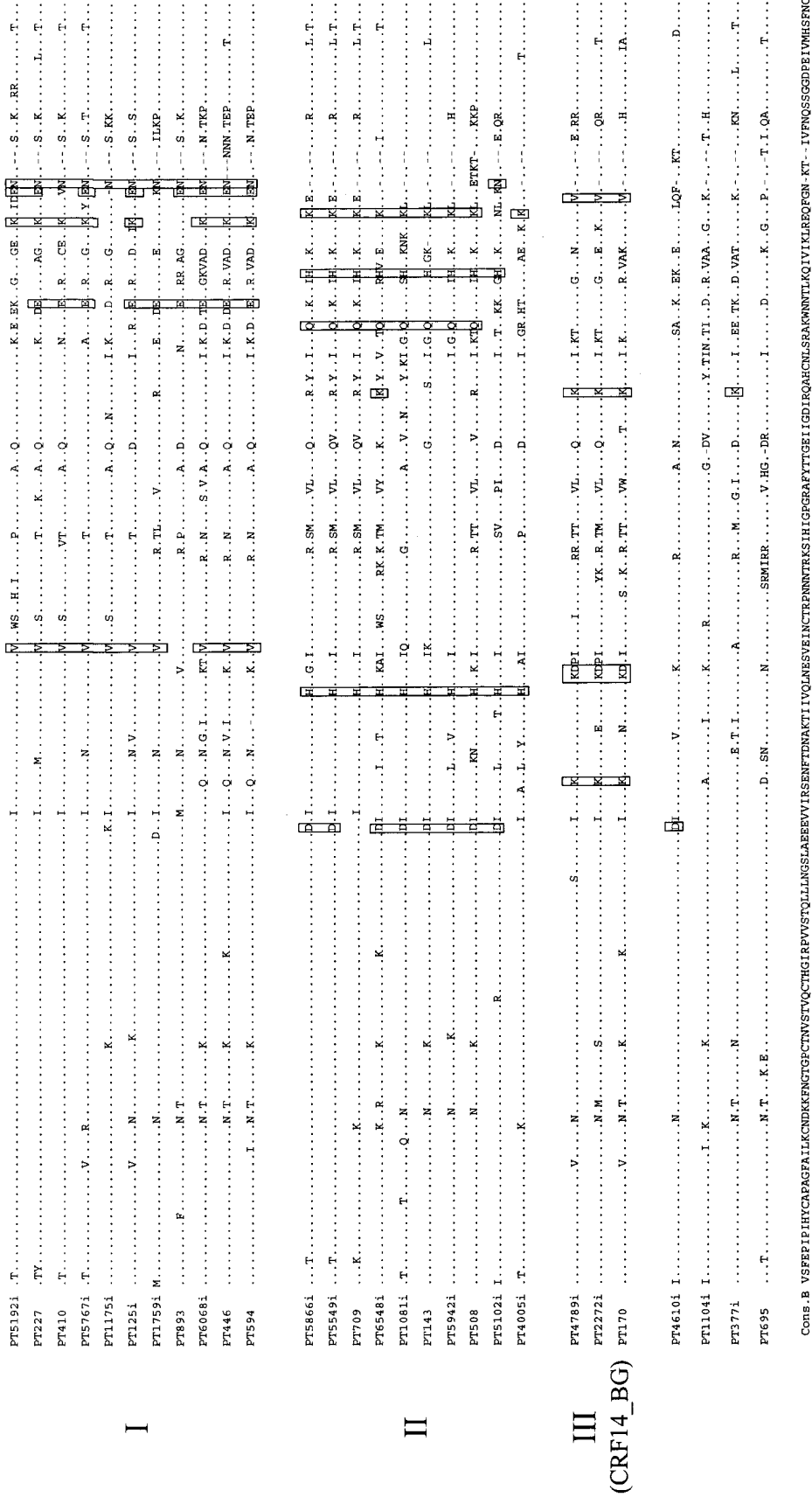
The phylogenetic tree in Fig. 1 shows that 29 subtype G viruses from Portuguese IDUs are assigned to a monophyletic cluster (bootstrap value of 961), enlarging the previously described G<sup>P</sup> group.<sup>2</sup> Only one G viral sequence (PT6123i) was found outside of the G<sup>P</sup> cluster, being placed among other G reference sequences. The genetic diversity of nucleotide sequences for subtype G viruses ranged from 2.2 to 23.4% (mean,  $11.6 \pm 3.4\%$ ). Within the G<sup>P</sup> subcluster the mean distance was  $11.0 \pm 2.6\%$  (ranging from 2.2 to 18.9%). Assuming a comparable evolution rate for different HIV-1 subtypes [(6.7  $\pm$

$2.1) \times 10^{-3}$  substitutions/site per year in V3<sup>13</sup>], G strains might have been introduced more than a decade ago among Portuguese IDUs. About two-thirds of the G<sup>P</sup> viruses previously described<sup>2</sup> were found in individuals of African origin with reported heterosexual risk of transmission, and the Portuguese IDUs may have been infected from this group.

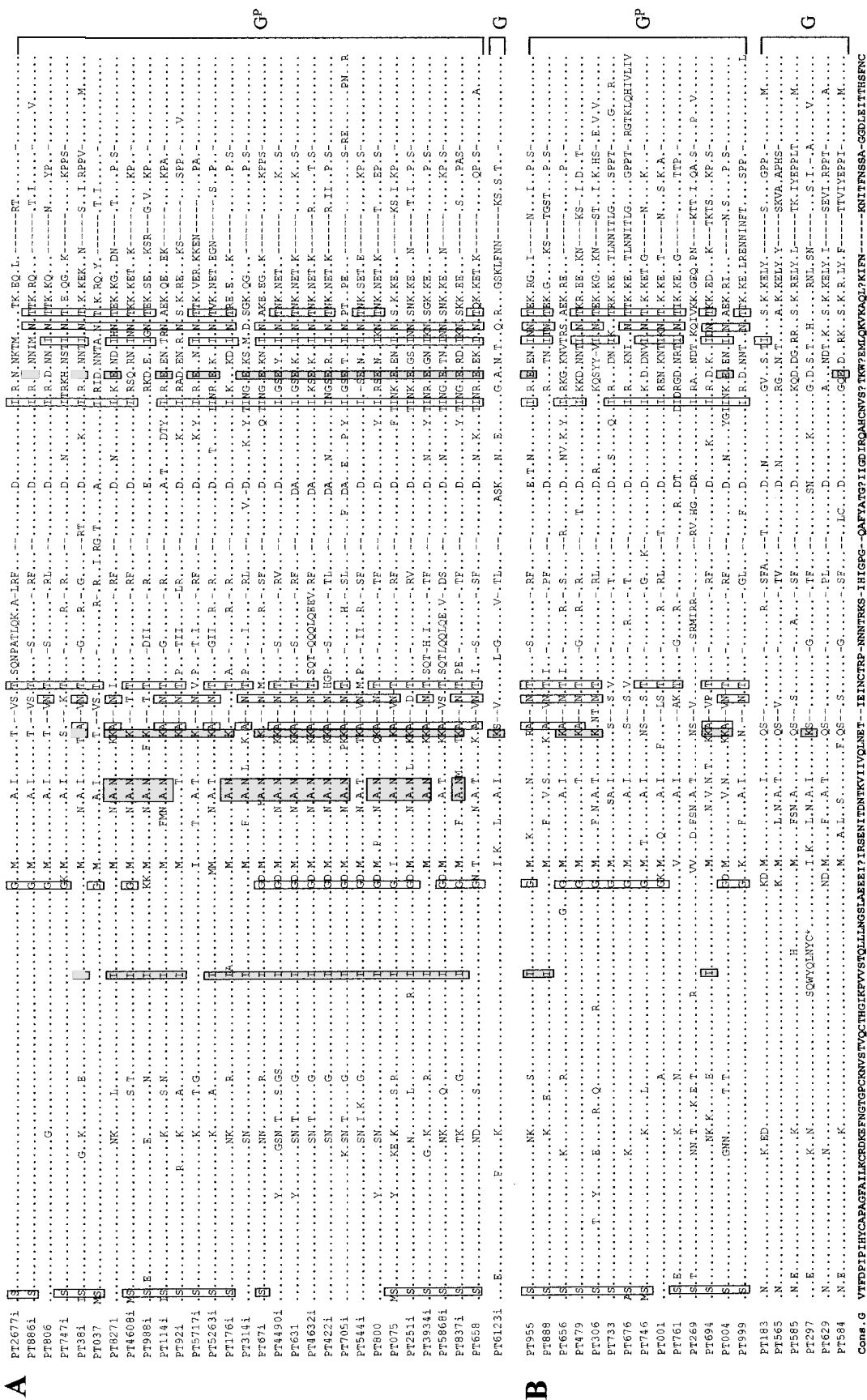
The predicted amino acid sequences of the *env* C2V3C3 region obtained for 28 subtype B strains, aligned with the B amino acid consensus, are presented in Fig. 2. All isolates but three have cysteines at both ends of the V3 termini. Two sequences (PT695 and PT1104i) have V3 loops with 34 amino acids, in opposition to the more common 35-amino acid V3 length, because of a deletion corresponding to amino acid position 320 in HXB2. At the crown of the V3 loop, 4 different tetrameric sequences were found but the GPGR motif is conserved (24 of 27). Subtype B amino acid sequences could be assigned to three distinct groups (I, II, and III in Fig. 2), with correspondent clusters I, II, and III in the phylogenetic tree of Fig. 1, on the basis of statistically significant differences (two-tailed Fisher exact test,  $p < 0.001$ ) in the frequency of nonsynonymous mutations (boxed in Fig. 2) between them. In addition, group I possesses one insertion (asparagine) at *env* codon 356 of HXB2 ( $\chi^2$  test with the Yates correction,  $p < 0.001$ ). These amino acid signatures are rarely found in published C2V3C3 sequences (HIV Sequence Database, <http://hiv-web.lanl.gov/>) and may reflect clonal dissemination of B strains from different point introductions.

Figure 3 shows the deduced amino acid sequences of the C2V3C3 region from subtype G viruses infecting the IDUs studied (Fig. 3A) aligned with subtype G amino acid sequences from individuals, originating from the same geographic area, who were reported to have been infected heterosexually (Fig. 3B) and previously studied by us.<sup>2</sup> Both risk groups include G<sup>P</sup> and non-G<sup>P</sup> sequences. Within the IDU subtype G group, the prevalent crown tetrapeptide motif is GPGQ (28 of 30), which is the most common motif found in non-B subtypes.<sup>14</sup> The majority of the V3 loops contain 35 amino acid residues, although three samples possess 36 because of single insertions at Env positions 300 (PT2677i and PT5868i) and 308 (PT4632i) of HXB2, and one (PT037) has 37 amino acid residues because of a 2-amino acid insertion at the V3 tip. Nine noncontiguous amino acids identified (boxed in Fig. 3) as characteristics of the G<sup>P</sup> sequences were underrepresented, or not present at all, in the other G viral sequences (two-tailed Fisher exact test,  $p < 0.01$ ) in both risk groups. Four of these amino acid signatures were already described<sup>2</sup> but the higher number of sequences analyzed in this study allowed us to establish new ones and also to define three silent substitutions (**TTG**, **CTA**, and **TTA**, HXB2 *env* codons 259, 261, and 369, respectively; two-tailed Fisher exact test,  $p < 0.001$ ). In addition, the amino acid motif AKN (Env positions 281–283), and the three nonsynonymous substitutions V255I, E290K, and K337E (boxed and shaded in Fig. 3) were defined as signatures ( $\chi^2$  test with the Yates correction;  $p < 0.01$ ) for G amino acid sequences from the IDU group and considered atypical in the G sequences from the heterosexual group. These signatures have so far an unknown biological significance. Alternatively, they may be due to a founder effect of neutral alleles.

In conclusion, the HIV-1 epidemic in this Portuguese IDU group is quite distinct from those previously described for other



**FIG. 2.** Alignments of deduced protein sequences of the *env* C2V3C3 region of HIV-1 subtype B strains amplified from Lisbon IDUs. Dots indicate identity to the consensus and dashes correspond to gaps introduced to maintain the correct alignment. The subtype B consensus sequence (Cons.B) has been compiled from the HIV Sequence Database (<http://hiv-web.lam1.gov/>). The positions found to correspond to statistically significant amino acid signatures are included in rectangles, and were used to define groups I-III (the same as those indicated in Fig. 1).



**FIG. 3.** Alignments of putative protein sequences of the *env* C2V3C3 region of Lisbon subtype G HIV-1 strains. Sequences obtained from IDUs and those from individuals who were reported to have been infected by the heterosexual route are grouped in (A) and (B), respectively. Dots indicate identity to the consensus, dashes correspond to gaps introduced to maintain the correct alignment, question marks refer to highly variable amino acid positions, and asterisks indicate the introduction of stop codons. Cons G represents the HIV-1 subtype G consensus sequence compiled from the HIV Sequence Database (<http://hiv-web.lanl.gov/>). The G<sup>p</sup> subgroup is described in a previous report.<sup>2</sup> Amino acid positions that represent signatures for IDU-derived sequences, when compared with heterosexually acquired HIV-1 sequences, are boxed and shaded in gray. Positions that represent signatures for the G<sup>p</sup> subcluster, when compared with G viruses, are boxed without shading. Statistical analysis was performed as stated in text.

IDUs in Europe.<sup>15</sup> At least 46.0% of the HIV-1 gene sequences analyzed are derived from non-B viruses and the viral genetic diversity observed is no longer determined by the inclusion of African subjects, because there is no association between viral subtype distribution and geographic origin of the seropositive individuals. About three-quarters of the non-B strains have subtype G sequences and approximately half of those are potential recombinant forms. Subtype G nonrecombinant and recombinant forms were reported for IDUs in Galicia, Spain, but associated with much lower prevalences.<sup>5</sup> A longstanding presence of HIV-1 G strains in Portugal is suggested by the high values of interindividual genetic distances, close to those obtained for B viruses, and by the variety of genetic forms found. This hypothesis will be confirmed by retrospective epidemiological studies.

### ACKNOWLEDGMENTS

Our special thanks to the Japan Health Sciences Foundation (Tokyo, Japan) for partially funding of this study. The following reagents were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: heteroduplex mobility analysis genetic subtyping kit (*gag*) from Drs. L. Heyndrickx, G. Van der Auwera, and G. van der Groen and the DAIDS, NIAID; heteroduplex mobility analysis genetic subtyping kit (*env*) from Drs. E. Delwart, B. Herring, J.I. Mullins, and the DAIDS, NIAID. We also thank the technical staff of Centro de Acolhimento de Alcântara, Rua de Cascais, Câmara Municipal de Lisboa and Gabinete de Reconversão do Casal Ventoso, Avenida de Ceuta, Lisboa, for their kind cooperation.

### REFERENCES

1. Anonymous: Global AIDS surveillance. Part II. *Wkly Epidemiol Rec* 2001;76:390-396.
2. Esteves A, Parreira R, Venenno T, *et al.*: Molecular epidemiology of HIV type 1 infection in Portugal: High prevalence of non-B subtypes. *AIDS Res Hum Retroviruses* 2002;18:313-325.
3. 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.
4. Page RD: TreeView: An application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 1996;12:357-358.
5. 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.
6. Delgado E, Thomson MM, Villahermosa ML, *et al.*: Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure. *J Acquir Immune Defic Syndr* 2002;29:536-543.
7. Liitsola K, Tashkinova I, Laukkanen T, *et al.*: HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad. *AIDS* 1998;12:1907-1919.
8. Lukashov VV, Huismans 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.
9. Lukashov VV, Kuiken CL, Vlahov D, Coutinho RA, and Goudsmit J: Evidence for HIV type 1 strains of U.S. intravenous drug users as founders of AIDS epidemic among intravenous drug users in northern Europe. *AIDS Res Hum Retroviruses* 1996;12:1179-1183.
10. Kuiken C, Thakallapalli R, Esklid A, and de Ronde A: Genetic analysis reveals epidemiologic patterns in the spread of human immunodeficiency virus. *Am J Epidemiol* 2000;152:814-822.
11. Adwan G, Papa A, Kouidou S, *et al.*: HIV type 1 sequences with GGC substitution in injecting drug users in Greece. *AIDS Res Hum Retroviruses* 1999;15:679-680.
12. Stoeckli TC, Steffen-Klopfstein I, Erb P, Brown TM, and Kalish ML: Molecular epidemiology of HIV-1 in Switzerland: Evidence for a silent mutation in the C2V3 region distinguishing intravenous drug users from homosexual men. *J Acquir Immune Defic Syndr* 2000;23:58-67.
13. Leitner T and Albert J: The molecular clock of HIV-1 unveiled through analysis of a known transmission history. *Proc Natl Acad Sci USA* 1999;96:10752-10757.
14. Vasil S, Thakallapalli R, Korber B, and Foley B: Global variation in the HIV-1 V3 region. In: *Human Retroviruses and AIDS* (Korber B, Kuiken C, Foley B, *et al.*, eds.). Los Alamos National Laboratory, Los Alamos, New Mexico, 1998, pp. III118-III265.
15. Op de Coul ELM, Prins M, Cornelissen M, *et al.*: Using phylogenetic analysis to trace HIV-1 migration among Western European injecting drug users seroconverting from 1984 to 1997. *AIDS* 2001;15:257-266.

Address reprint requests to:

Aida Esteves  
 Unidade de Virologia  
 Instituto de Higiene e Medicina Tropical  
 Rua da Junqueira, 96  
 P-1349-008 Lisbon, Portugal

E-mail: aidaesteves@ihmt.url.pt