

Sequence Note

Near Full-Length Genome Analysis of HIV Type 1 CRF02_AG, Subtype C and CRF02_AG Subtype G Recombinants

WOUTER JANSSENS,^{1,2} MIKA O. SALMINEN,³ TIINA LAUKKANEN,³ LEO HEYNDRICKX,¹
GERT VAN DER AUWERA,¹ ROBERT COLEBUNDERS,⁴ FRANCINE E. McCUTCHAN,⁵ and
GUIDO VAN DER GROEN¹

ABSTRACT

HIV-1 CRF02_AG strains are prevalent in west and west-central Africa, suggesting a longstanding presence of these subtype A/G recombinants in the global epidemic. Cocirculation of CRF02_AG strains with other group M subtypes may give rise to HIV-1 recombinants constituting a mosaic genome comprising fragments of three different subtypes. We report on the genetic analysis of the near-full-length genomes of such recombinants (VI1035 and VI1197) as well as CRF02_AG strains in Belgian individuals. VI1035 and VI1197 may be the result of successful “second-generation” recombinations of HIV-1 strains CRF02_AG with, respectively, subtype C (VI1035) and G (VI1197) strains in a dually infected individual.

THE RECOGNITION that intersubtype recombination is an important source of HIV-1 variation has led to new HIV-1 classification criteria. A subtype is now defined as a monophyletic cluster of three epidemiologically unlinked variants. The monophyletic clustering must be maintained regardless of the subgenomic segment analyzed to fulfill the criterion for a separate subtype. HIV-1 group M is currently subdivided in subtypes A, B, C, D, F, G, H, J, and K. Following the same lines, circulating recombinant forms (CRFs) are defined as intersubtype recombinants for which at least three epidemiologically unlinked variants are monophyletic and thus have an identical mosaic genome structure. So far CRF01_AE, CRF02_AG, CRF03_AB, and CRF04_cpx fulfill these criteria.¹ This classification forms a framework to identify new subtypes and intersubtype recombinants that are an important source of HIV-1 genetic variation. Whereas CRF03_AB and CRF04_cpx apparently still have a relatively limited spread, CRF01_AE and CRF02_AG have been well established into different popula-

tions most likely as a result of a founder effect. CRF01_AE is common in the Central African Republic, Thailand, and other Asian countries.² Near-full-length CRF02_AG recombinants were identified in Nigeria, Djibouti, and Côte d'Ivoire.³ Based on genetic analysis of small genome fragments, a high prevalence of CRF02_AG was documented in west and west-central Africa suggesting a longstanding presence of these recombinants in the global epidemic. Indeed, reanalysis of *env* subtype A strains documented before the recognition of CRF02_AG, as well as analysis of newly documented samples indicate the epidemiological relevance of this CRF. Reanalysis of partial *env* sequences from Côte d'Ivoire and Cameroon, initially documented as subtype A, indicates that strains of HIV-1 similar to CRF02_AG are the most common form of subtype A in west and west-central Africa.³ We demonstrated by reanalysis of *env* sequences encoding the region C2V3 till the start of gp41 (900 base pairs) of subtype A samples from Côte d'Ivoire and Cameroon, in combination with *gag* HMA that CRF02_AG

¹Department of Microbiology, Institute of Tropical Medicine, B-2000 Antwerp, Belgium.

²The Flanders Interuniversity Institute for Biotechnology, VIB, B-9052 Ghent, Belgium.

³HIV Laboratory, Department of Infectious Disease Epidemiology, National Public Health Institute, FIN-00300 Helsinki, Finland.

⁴Department of Clinical Sciences, Institute of Tropical Medicine, B-2000 Antwerp, Belgium.

⁵Henry M. Jackson Foundation for the Advancement of Military Medicine, Rockville, Maryland 20850.

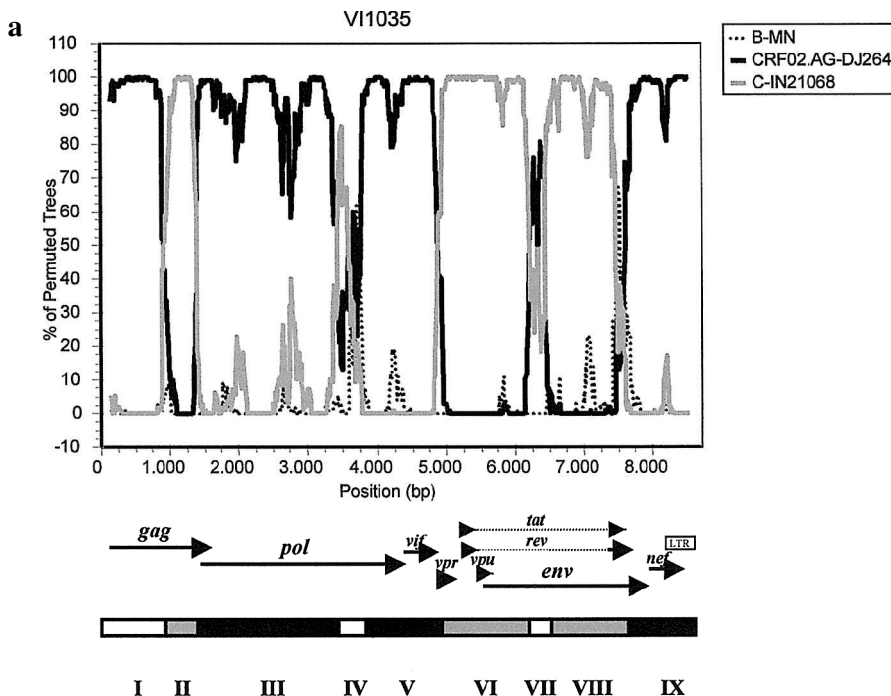


FIG. 1. (a) Bootscanning analysis of the complete genome of VI1035. Genome alignment positions and bootstrap values (%) are depicted in the *x*- and *y*-axis, respectively. VI1035 is compared with representatives of subtype B (MN; outgroup), C (IN21068), and CRF02.AG (DJ264). White regions I, IV, and VII cannot be assigned to a particular subtype or CRF by phylogenetic analysis. (b) Phylogenetic classification of genome segments based on an alignment of HIV-1. The phylogenetic position of VI1035 is indicated in *italic*; 2000 bootstrap samples were analyzed. Bootstrap values are given in percentages at the internodes if they exceed the 70% level. The tree is rooted arbitrarily between subtypes.

strains were already prevalent in these countries in the early 1990s.⁴ We recently documented by *gag* HMA⁴ that CRF02.AG is prevalent in Yaoundé (Cameroon) and Cotonou (Bénin), comprising about 40% of the documented samples. Another recent study on *env* C2V3 phylogenetic analysis of samples from the northern part of Cameroon documented one-third of the analyzed samples as CRF02.AG variants.⁵ The longstanding presence of CRF02.AG strains in regions where other group M subtypes cocirculate may give rise to a next generation of HIV-1 recombinants constituting a mosaic genome comprising fragments of three different subtypes. Here we report on the genetic analysis of the near-full-length genomes of such recombinants (VI1035 and VI1197) as well as CRF02.AG strains in Belgian individuals.

The Belgian individual infected with VI1035 (isolated June 30, 1993) was probably infected by his female Kenyan partner from the border region between Kenya and Tanzania. He later also had multiple heterosexual contacts in other parts of Africa and Asia. The Belgian individual infected with VI1197 (isolated February 1, 1994) was infected by blood transfusion in Belgium in June 1989. VI1144, a follow-up sample of VI1035, was previously classified as subtype C by *env* HMA,⁶ but could not be classified by C2V3 phylogeny.⁷ VI1197 was previously classified as *env* subtype G by *env* HMA⁶ and phylogenetic analysis of a 900 base pair (bp) fragment encoding C2V3 until the start of gp41.⁸

Near-full-length genomes of VI1035 and VI1197 were amplified by polymerase chain reaction (PCR) from DNA isolated

from cocultured peripheral blood mononuclear cells (PBMC), cloned and sequenced as described previously.^{9,10} The lengths of the clones of VI1035 and VI1197 were 8945 and 8914 bp, respectively. All reading frames were open and of complete length. The nucleotide sequence data were deposited in the EMBL, GenBank, and DDBJ Nucleotide Sequence Databases under the following accession numbers: AJ276595 (VI1035) and AJ276596 (VI1197).

VI1035 and VI1197 were compared with a panel of 24 near-full-length HIV-1 including group M subtypes A (UG037, Q23), B (MN, SF2), C (ETH2220, BR025), D (ELI, UG114), F (VI850, BR020), G (HH8793, SE6165), H (VI997, 90CF056.1), J (SE9280, SE9173), K (MP535, EQTB11), and CRF01.AE (CM240, CF402), CRF02.AG (IbNG, DJ263), and CRF04.cpx (CY032.3, PVMY) reference sequences in order to verify their recombinant status. Near full-length HIV-1 sequences were aligned automatically using CLUSTAL W,¹¹ and the resulting alignment was refined manually with the alignment editor DCSE.¹² Regions that could not be aligned unambiguously due to excessive length or sequence variation were omitted from the analysis. Similarity plot and bootscanning were performed using a sliding window of 300 base pairs (bp) advanced in 10-bp increments using SimPlot software.¹³ Bootscanning phylogeny involved generation of 1000 replicates, distance calculation by Kimura (2-parameter), transition/transversion ratio 2.0, and tree construction by neighbor joining. Plots of similarity were generated for identification of reference sequences having the highest similarity to (parts of)

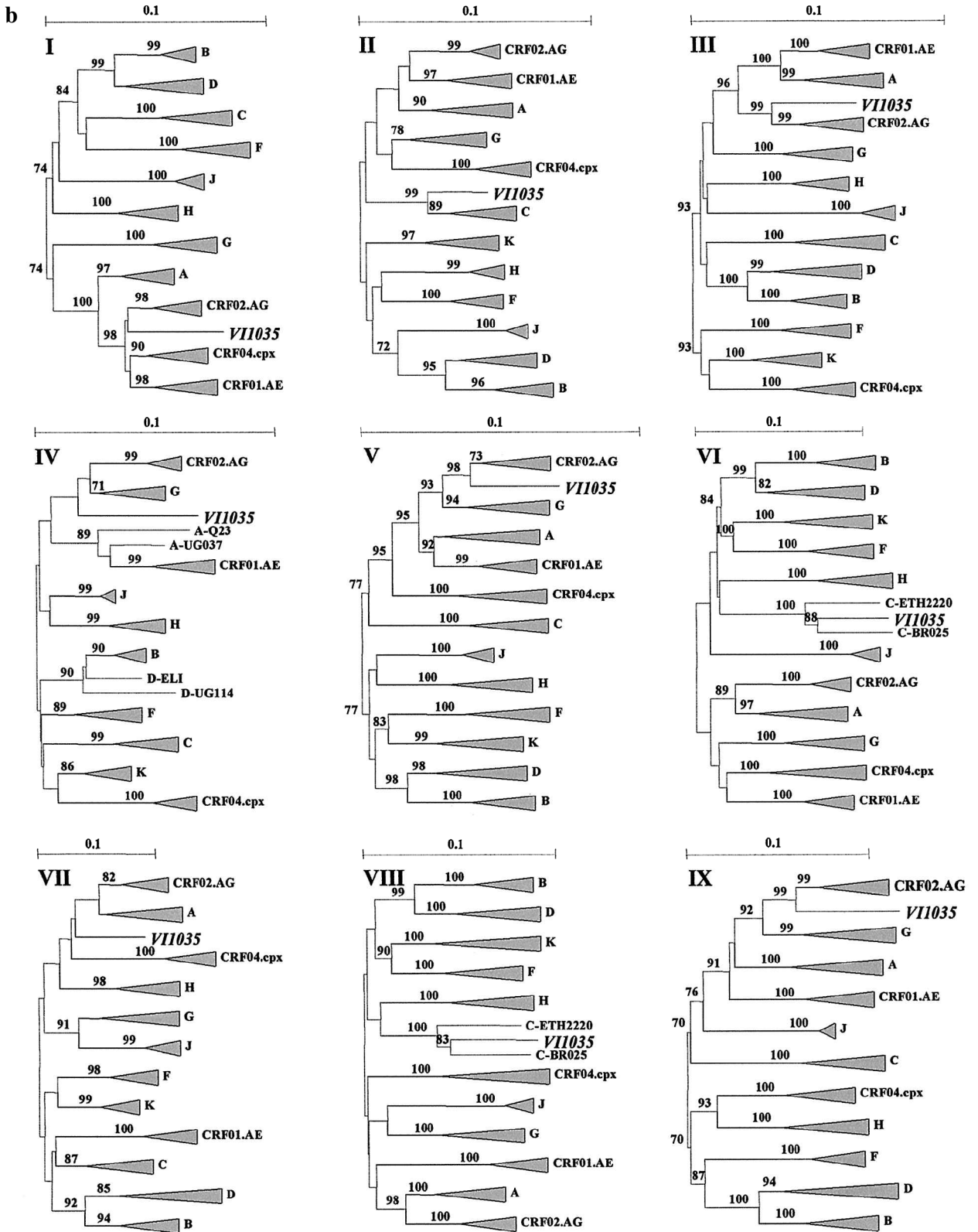


FIG. 1. Continued.

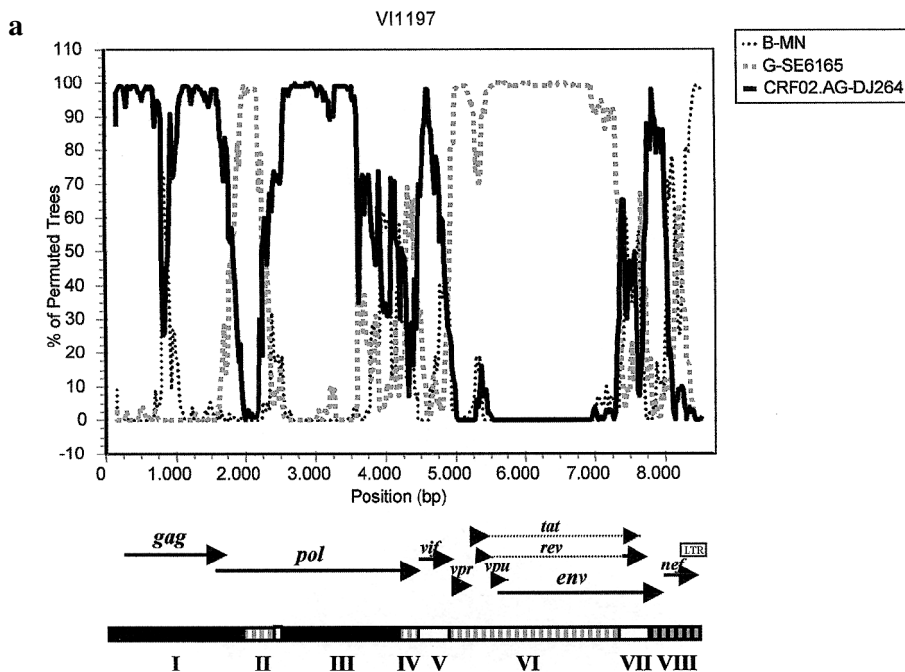


FIG. 2. (a) Bootscanning analysis of the complete genome of VII197. Genome alignment positions and bootstrap values (%) are depicted in the *x*- and *y*-axis, respectively. VII197 is compared with representatives of subtype B (MN; outgroup), G (SE6165), and CRF02.AG (DJ264). White regions V and VII cannot be assigned to a particular subtype or CRF by phylogenetic analysis. (b) Phylogenetic classification of genome segments based on an alignment of HIV-1. The phylogenetic position of VII197 is indicated in *italic*; 2000 bootstrap samples were analyzed. Bootstrap values are given in percentages at the internodes if they exceeded the 70% level. The tree is rooted arbitrarily between subtypes.

VII1035 and VII197 (results not shown) and used for bootscanning analysis. Similarity plot analysis reveals mosaic structures including genome segments that are CRF02.AG like and subtype C (VII1035) or G (VII197) segments; the graph of bootscanning analysis identifies the breakpoints (Figs. 1a and 2a). To confirm these results, the software package TREECON¹⁴ was used for distance calculations (Kimura), tree construction (neighbor joining), and bootstrap analysis (Figs. 1b and 2b).

The genome structure of VII1035 was as follows (Figs. 1a and b, and 3). A segment comprising the *gag* encoding region up to half of p24 clustered with subtype A, CRF01.AE, CRF02.AG, and CRF04.cpx [I; alignment position (AP) 1–875; genome position (GP) 1–886]. A BLAST search (15; <http://hiv-web.lanl.gov/>) did not reveal any significant relationship of (parts of) this fragment to subtype A, CRF01.AE, CRF02.AG, or CRF04.cpx. The second half of the *gag* p24 region clusters with subtype C (II; AP 876–1375; GP 887–1383). Where the *pol* gene starts VII1035 turns to CRF02.AG for a fragment encoding protease, reverse transcriptase, and most of RNase H (III; AP 1376–3375; GP 1384–3396). A segment of the *pol* region encoding the remaining part of the RNase H as well as part of the integrase remains unclassified (IV; AP 3376–3775; GP 3397–3795). The remainder of the *pol* gene and *vif* cluster with CRF02.AG (V; AP 3776–4875; GP 3796–4894). A region comprising *vpr*, *tat* (exon 1), *rev* (exon 1), *vpu*, and an *env* gp120 segment encoding V1 and V2 is subtype C (VI; AP

4876–6150; GP 4895–6315). The *env* C2V3 region remains unclassified (VII; AP 6151–6450; GP 6316–6585). A fragment encoding the remaining part of *env* gp120, the external part and the membrane spanning domain of gp41, is subtype C (VIII; AP 6451–7525; GP 6586–7758). The *env* encoding internal part of gp41, the second exons of *tat* and *rev*, *nef*, and part of LTR cluster with CRF02.AG (IX; AP 7526–8689; GP 7759–8945).

The genome structure of VII197 was as follows (Figs. 2a and b, and 3). A segment comprising the *gag* gene and the protease encoding part of the *pol* gene clustered with CRF02.AG (I; AP 1–1900; GP 1–1707). Part of the *pol* reverse transcriptase encoding region is subtype G (II; AP 1901–2300; GP 1708–2306). A *pol* gene segment encoding part of reverse transcriptase, RNase H, and part of integrase clusters with CRF02.AG (III; AP 2401–4100; GP 2407–4102). The remainder of the *pol* gene of VII197 clusters with subtype G (IV; AP 4101–4400; GP 4103–4401). Most of the *vif* gene remains unclassified (V; AP 4401–4800; GP 4402–4801). A segment including the terminal part of the *vif* gene, *vpr*, *tat* (exon 1), *rev* (exon 1), *vpu*, *env* gp120, and the external part of gp41 encoding regions is classified as subtype G (VI; AP 4801–7300; GP 4802–7489). A segment encoding the *env* membrane spanning domain, part of the internal part of gp41, the second exon of *tat*, and part of the second exon of *rev*, remains unclassified (VII; AP 7301–7700; GP 7490–7884). The remaining part of the *env* gp41 and the *rev* second exon encoding regions, *nef*, and part of LTR cluster with subtype G and CRF02.AG (VIII; AP 7701–8689; GP 7885–8914).

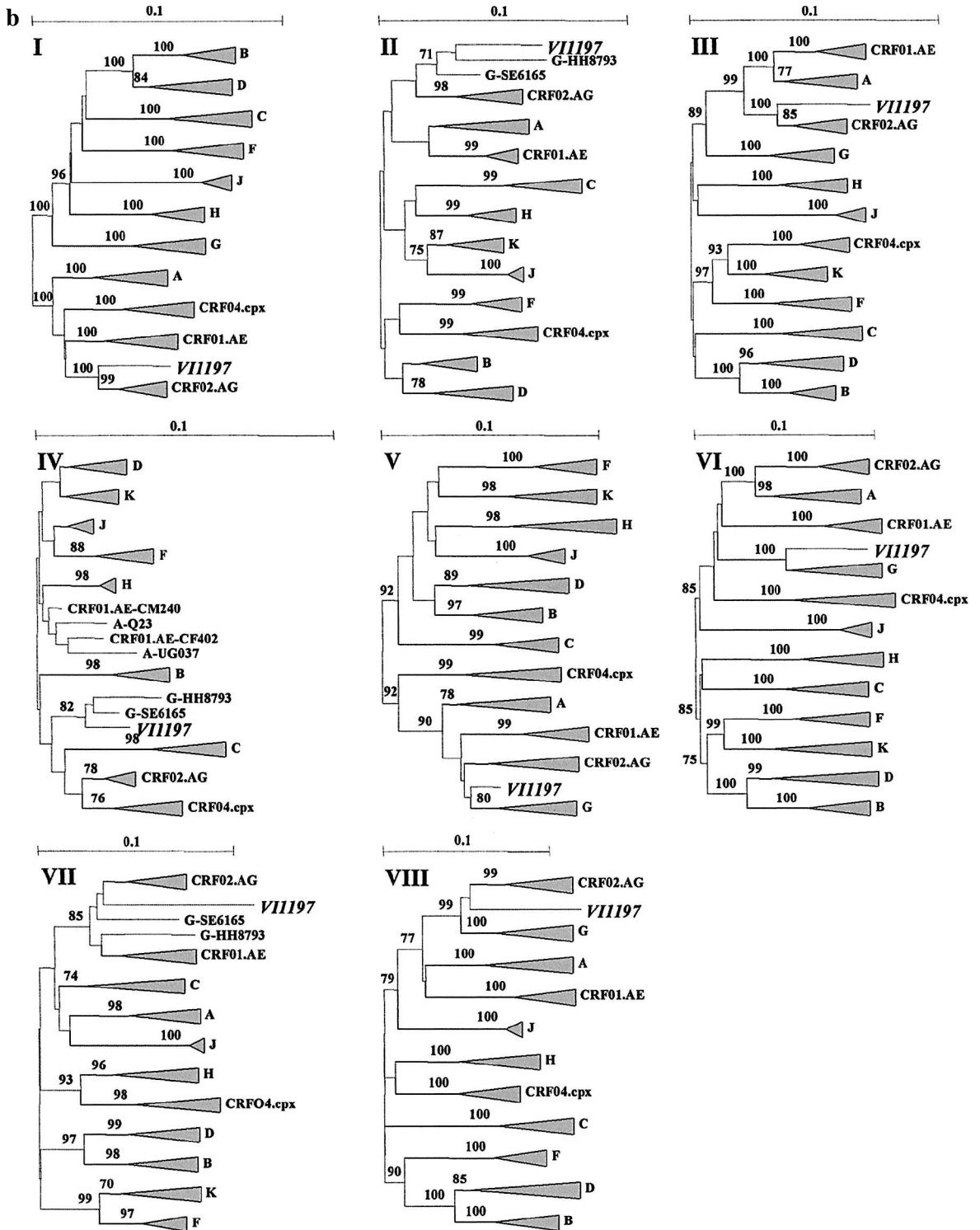


FIG. 2. Continued.

The long-standing presence of non-B strains in HIV-1-seropositive Belgian individuals was documented earlier.¹³ Generally these infections were linked to regions in the world where these strains are prevalent. To verify to what extent CRF02.AG variants were introduced into the Belgian popula-

tion, we screened previously documented HIV-1 *env* subtype A-infected individuals¹⁴ for actual CRF02.AG infections by phylogenetic analysis including CRF02.AG references, as well as by *gag* HMA.⁶ Four of 29 “A” variants, introduced to the Belgian population 8 to 10 years ago, were classified as

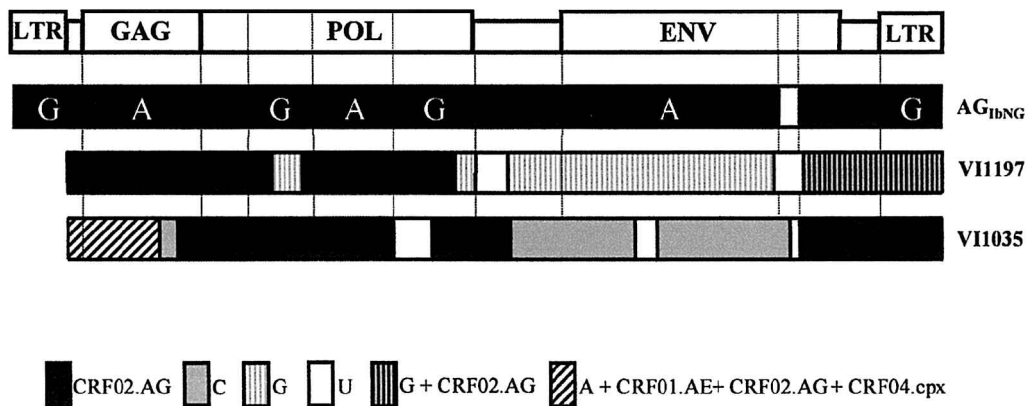


FIG. 3. Diagram of the genotypic structure of CRF02.AG, VII197, and VII035. Vertical lines indicate subsequent genome fragments determining the recombinant genome structure of CRF02.AG was documented previously.⁴ The segments that are unlabeled in CRF02.AG are unassigned to any subtype, but common to CRF02.AG strains. Black regions indicate a monophyletic relationship between CRF02.AG, VII035, and VII197. White-gray hatched regions in VII197 cluster with representatives of subtype G. Gray regions in VII035 cluster with representatives of subtype C. The gray-black hatched pattern region in VII197 clusters with subtype G and CRF02.AG. The white-black hatched region in VII035 clusters with subtype A, CRF01.AE, CRF02.AG, and CRF04.cpx. White regions are unclassified (U).

CRF02.AG by *gag* HMA, of which three phylogenetically clustered with *env* C2V3 CRF02.AG references supported by bootstrap analysis (results not shown; Table 1). These findings indicate a longstanding presence of HIV-1 CRF02.AG strains in Belgium.

It was previously recognized that the frequency of documenting HIV-1 recombinants in seropositive individuals correlated with cocirculation of multiple HIV-1 subtypes in populations in particular geographic regions. Since HIV-1 CRF02.AG strains are prevalent in west and west-central Africa, it may not be surprising that CRF02.AG variants contributed to “second-generation” recombinants comprising fragments of three different subtypes, that are transmitted into the population. In addition to the mosaic CRF02.AG-C (VII035) and CRF02.AG-G (VII197) genomes documented in this study, we previously identified one isolate of *gag/env* A/CRF02.AG, CRF02.AG/F, and CRF02.AG/G recombinants in Côte d’Ivoire,¹⁰ Cameroon,⁶ and Bénin,⁶ respectively.

VII035 and VII197 may be the result of a successful re-

combination of HIV-1 strains of CRF02.AG with, respectively, subtype C (VII035) and G (VII197) strains in a dually infected individual. The evolution toward “second-generation” recombinants contributes to the unpredictability of evolutionary routes in the further evolution of HIV-1.

ACKNOWLEDGMENTS

This work was supported by the Fonds voor Wetenschappelijk Onderzoek, Brussels, Belgium (Grant G.0134.97); the Flanders Interuniversity Institute for Biotechnology (VIB), Ghent, Belgium; the Human Science Foundation (Tokyo, Japan); and in part by a cooperative agreement between the Henry M. Jackson Foundation for the Advancement of Military Medicine and the United States Army Medical Command. (The views and opinions expressed herein do not necessarily reflect those of the U.S. Army or the Department of Defense.)

TABLE 1. EPIDEMIOLOGICAL DATA OF HIV-1 SEROPOSITIVE INDIVIDUALS

Isolate number	Source of infection	Time of infection	Risk behavior	Genetic classification		
				<i>gag</i> ^a	<i>env</i> ^b	Genome
VI712	DRC	May 1992	Heterosexual	CRF02.AG	CRF02.AG	—
VI860	Nigeria	Between December 1989 and December 1992	Heterosexual	CRF02.AG	CRF02.AG	—
VI1007	Côte d’Ivoire	Before 1990	Heterosexual	CRF02.AG	A+CRF02.AG	—
VI1035	Africa and/or Asia	Before 1993	Heterosexual	C	U	CRF02.AG-C
VI1144				C	U	—
VI1197	?	?	Blood transfusion	CRF02.AG	G	CRF02.AG-G
VI1320	West-Africa	May 1990	Heterosexual	CRF02.AG	CRF02.AG	—

^aThe *gag* region as determined by *gag* HMA encodes from amino acid 132 of p24 until amino acid 40 of p7 (according to HIV-1 ELI; accession number K03454).

^bThe *env* C2V3 region as determined by sequencing and phylogenetic analysis.

REFERENCES

1. Robertson DL, Anderson JP, Bradac JA, et al.: HIV-1 nomenclature proposal—A reference guide to HIV-1 classification. HIV Sequence Database 2000. <http://hiv-web.lanl.gov/>
2. Gao F, Robertson DL, Morrison SG, et al.: The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J Virol* 1996;70:7013–7029.
3. Carr JK, Laukkanen T, Salminen MO, et al.: Characterization of subtype A HIV-1 from Africa by full genome sequencing. *AIDS* 1999;13:1819–1826.
4. Heyndrickx L, Janssens W, Zekeng L, et al.: Simplified strategy for detection of recombinant HIV-1 group M isolates by gag/env heteroduplex mobility assay. *J Virol* 2000;74:363–370.
5. Mboudjeka I, Zekeng L, Takehisa J, et al.: HIV type 1 genetic variability in the northern part of Cameroon. *AIDS Res Hum Retroviruses* 1999;15:951–956.
6. Fransen K, Buvé A, Nkengasong JN, Laga M, and van der Groen G: Longstanding presence in Belgians of multiple non-B HIV-1 subtypes. (letter). *Lancet* 1996;347:1403.
7. Heyndrickx L, Janssens W, Coopens S, et al.: HIV-1 C2V3 env diversity among Belgian individuals. *AIDS Res Hum Retroviruses* 1998;14:1291–1296.
8. Van de Peer Y, Janssens W, Heyndrickx L, Fransen K, van der Groen G, and De Wachter R: Phylogenetic analysis of the env gene of HIV-1 isolates taking into account individual nucleotide substitution rates. *AIDS* 1996;10:1485–1494.
9. Salminen MO, Koch C, Sanders-Buell E, et al.: Recovery of virtually full-length HIV-1 provirus of diverse subtypes from primary virus cultures using the polymerase chain reaction. *Virology* 1995;213:80–86.
10. Laukkanen T, Albert J, Liitsola K, et al.: Virtually full-length sequences of HIV-1 subtype J reference strains. *AIDS Res Hum Retroviruses* 1999;15:293–297.
11. Thompson JD, Higgins DG, and Gibson TJ: CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680.
12. De Rijk, P, and De Wachter R: DCSE, an interactive tool for sequence alignment and secondary structure research. *Comput Applic Biosci* 1993;9:735–740.
13. Ray SC: SimPlot for Windows (version 1.6). Baltimore, MD. Distributed by author. 1998. (<http://www.med.jhu.edu/deptmed/sray/download>).
14. Van de Peer Y, and De Wachter R: TREECON for Windows: A software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput Applic Biosci* 1994;10:569–570.
15. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, and Lipman DJ: Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.

Address reprint requests to:

Wouter Janssens
Department of Microbiology
Institute of Tropical Medicine
Nationalestraat 155
2000 Antwerp, Belgium

E-mail: wjanssens@itg.be