

Identification of 3 Phylogenetically Related HIV-1 BG Intersubtype Circulating Recombinant Forms in Cuba

María Sierra, PhD,* Michael M. Thomson, MD, PhD,* David Posada, PhD,† Lissette Pérez, BSc,‡ Carlos Aragonés, MSc,‡ Zoila González, BSc,‡ Jorge Pérez, MD, PhD,‡ Gema Casado, PhD,* and Rafael Nájera, MD, PhD*

Summary: BG intersubtype recombinants represented 11.6% of HIV-1 isolates in a recent survey in Cuba based on *pol* sequences, most of them forming a single clade further subdivided into 3 subclades. Here, we analyze 8 near full-length genomes and 1 *gag-pol* sequence from epidemiologically unlinked Cuban BG recombinants from these 3 subclades (3 from each). Near full-length sequences were also obtained from 3 subtype G and 2 subtype B Cuban viruses. Phylogenetic relationships were estimated via maximum likelihood, and mosaic structures of the recombinants were inferred with the bootscanning, MaxChi, Genconv, and GARD methods. For the near full-length genomes, all recombinants formed a strongly supported clade further subdivided into the same subclades previously defined in *pol*. Mosaic structures were identical within each subclade and different among subclades, although 5 breakpoints were coincident among all recombinants. Individual phylogenetic trees for non-recombinant fragments (concatenated B and G subtype segments) indicated a common ancestry for the parental viruses and their relationships to local subtype B and G strains. These results allow us to identify 3 new BG intersubtype circulating recombinant forms in Cuba derived from a common recombinant ancestor, which originated from B and G subtype parental strains circulating in Cuba.

Key Words: BG intersubtype recombinant, circulating recombinant form, Cuba, HIV-1, subtype B, subtype G

(*J Acquir Immune Defic Syndr* 2007;45:151–160)

Received for publication December 22, 2006; accepted February 7, 2007.

From the *Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda (Madrid), Spain; †Department of Biochemistry, Genetics, and Immunology, University of Vigo, Vigo (Pontevedra), Spain; and ‡Instituto de Medicina Tropical “Pedro Kourí,” La Habana, Cuba.

M. Sierra and M. M. Thomson contributed equally to this article.

Funded in part by grant 36205/01 from the Fondo para la Investigación y Prevención del SIDA en España (FIPSE), Spain. L. Pérez was supported by grant A40325 from the United Nations Program on HIV/AIDS.

Presented in part at the XV International AIDS Conference, Bangkok, Thailand, July 11–16, 2004, and at the 12th Conference on Retroviruses and Opportunistic Infections, Boston, MA, February 22–25, 2005.

GenBank accession numbers: AY586542–AY586545, AY586547–AY586549, DQ020274, AY900571–AY900576.

Reprints: Rafael Nájera, MD, PhD, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Ctra. Majadahonda-Pozuelo, Km. 2, 28220 Majadahonda (Madrid), Spain (e-mail: rafael.najera@isciii.es).

Copyright © 2007 by Lippincott Williams & Wilkins

HIV-1 exhibits a high degree of genetic diversity derived from rapid replication and elevated mutation and recombination rates. The recognition of recombination as a major mechanism for the generation of viral diversity in the HIV-1 pandemic is increasing in parallel to the growth in number and length of viral sequences available as well as with the expanding geographic coverage of epidemiologic molecular studies, which show that recombinant forms are frequently generated in areas in which diverse HIV-1 clades circulate in the same population.^{1,2} Some of the recombinants generated in such areas have become circulating recombinant forms (CRFs),³ several of which have propagated to wide epidemic extents. So far, near full-length genome sequences of 34 potential CRFs have been submitted to the Los Alamos HIV Sequence Database,⁴ and 24 have been reported in the literature (CRFs 1–16, 18, 19, 21, 28, 29, 31, 32, and 33).

We previously reported the presence of an unusually high HIV-1 diversity in Cuba,^{5,6} which contrasts sharply with the almost uniform predominance of subtype B elsewhere in the Caribbean area. Among the genetic forms circulating in Cuba, there are 2 CRFs of central African origin (CRF18_cpx⁷ and CRF19_cpx⁸) and subtypes B and G. In partial *pol* sequences from samples collected in 2003, 17.5% viruses were recombinant among different HIV-1 clades circulating in Cuba.⁶ Among these, BG recombinant viruses were predominant, representing 11.6% of the viruses analyzed. These recombinants were identified predominantly among men who have sex with men (MSM) in Havana City, among whom they have expanded recently, as inferred from recent dates of HIV-1 diagnoses. BG recombinants have propagated rapidly, reaching 31.4% of HIV-1 infections among MSM in Havana City diagnosed in 2003. Most of the BG recombinants formed 3 phylogenetically related clades in *pol*, corresponding to 3 different recombinant structures.

Here, we analyze near full-length genome sequences of BG recombinant, subtype G, and subtype B viruses from Cuba. We show that the BG recombinants of Cuba represent 3 new CRFs originating from a common recombinant ancestor and ultimately derived from subtype G and B parental strains circulating in Cuba.

METHODS

BG recombinant viruses were unambiguously identified by bootscan analyses of partial *pol* segments in 52 Cuban HIV-1-infected individuals in samples collected in 1999 (n = 3) and

2003 ($n = 49$).^{5,6} Samples were collected in Cuba, except for 1 (R77) obtained from a Cuban residing in Spain. In phylogenetic trees of a *pol* segment (protease and 0.9 kilobases [kb] of reverse transcriptase; Fig. 1), 50 of the recombinants grouped in a monophyletic group or clade, which comprised 3 subclades, each consisting of 23 (subclade I), 20 (subclade II), and 6 (subclade III) viruses, respectively, plus 1 virus branching outside the 3 subclades.

Near full-length genome sequences were obtained from 3 viruses of subclade I (R77, Cu103, and CB_134), 3 viruses of subclade II (CB_378, CB_471, and CB_619), and 2 viruses of subclade III (CB_118 and CB_347), and the *gag-pol* sequence was obtained from 1 virus of subclade III (CB_418). According to partial *pol* and *env* sequences, 3 subtype G (Cu74, Cu85, and Cu87) and 2 subtype B (Cu19 and Cu43) viruses from Cuba were also amplified and sequenced in near full-length genomes; these viruses were selected for further analysis because they clustered in *pol* with segments of the Cuban BG recombinants (results not shown); therefore, they were our best candidates to represent the parental strains. Questionnaires carried out for contact tracing of HIV-1 infection did not reveal epidemiologic relationships among any of the individuals harboring the viruses analyzed here. Epidemiologic and demographic data are shown in Table 1.

Plasma RNA (in BG recombinant samples, except R77) or proviral DNA from peripheral blood mononuclear cells' lysates (in the remaining samples) was used for amplification. RNA was extracted from 1 mL of plasma using the reagents provided in the Nuclisens kit (Biomérieux, Marcy l'Etoile, France) according to the method by Boom et al.⁹ Near full-length genomes (~9 kb) were amplified by nested polymerase chain reaction (PCR); coupled with prior reverse transcription in RNA samples) in 4 overlapping fragments, as previously described;¹⁰⁻¹² *gag-pol* was amplified in 1 sample of subclade III (CB_418) by reverse transcription-nested PCR in 2 overlapping fragments. Enzymatically purified PCR products were directly sequenced by using the ABI Prism BigDye Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, CA). Products were detected in an ABI 3730 DNA Analyzer (Applied Biosystems) automated capillary sequencer.

Sequences were aligned using MAFFT version 5.3.¹³ Phylogenetic relationships were estimated by maximum likelihood with TREEFINDER,¹⁴ using the best-fit model of nucleotide substitution selected under the Akaike Information Criterion in MODELTEST version 3.6.¹⁵ Reliability of tree topologies was assessed by bootstrapping using 100 replicates. Recombinant structures were inferred by bootscanning¹⁶ using Simplot version 3.5.1.¹⁷ To locate more precisely the positions of breakpoints, the programs Genconv,¹⁸ MaxChi,¹⁹ and GARD²⁰ were used. To test the confidence of tree topologies for short (<200 nucleotides) interbreakpoint segments, the Shimodaira-Hasegawa (SH) test²¹ was applied using PAUP version 4b10.²²

RESULTS

For near full-length genomes, all 8 Cuban BG recombinants formed a single clade supported by a 100%

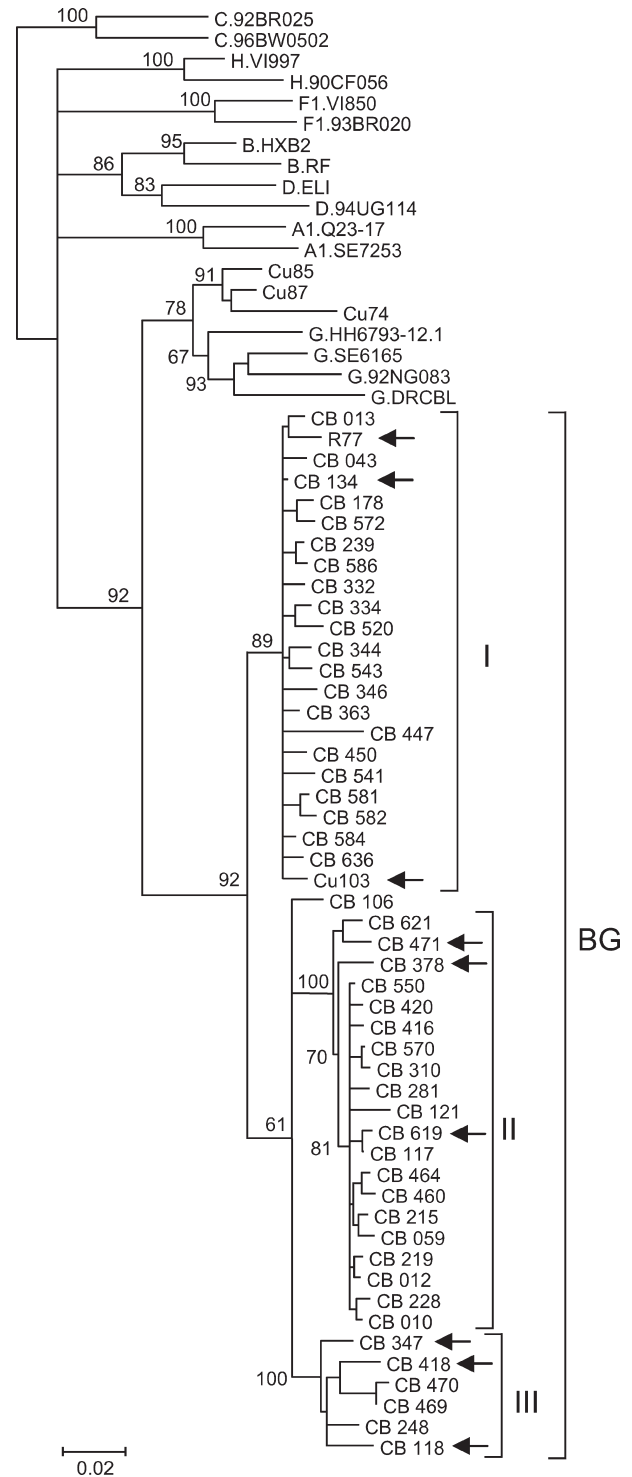


FIGURE 1. Maximum likelihood phylogenetic tree of protease-reverse transcriptase sequences of BG recombinants from Cuba. Trees were constructed with TREEFINDER,¹¹ using the best-fit model of nucleotide substitution selected under the Akaike Information Criterion in MODELTEST version 3.6,¹² with assessment of the reliability of tree topologies by bootstrapping using 100 replicates. Only bootstrap values $\geq 50\%$ are shown. Arrows signal BG recombinant viruses that were selected for further sequence analyses.

TABLE 1. Data of Analyzed Samples

Sample Identification	Year of Sample Collection	Gender	Risk Category	Country of Residence	Country of Origin	Year of HIV-1 Diagnosis	Subtype*
Cu19	1999	F	Hetero	Cuba	Cuba	na	B
Cu43	1999	M	MSM	Cuba	Cuba	1996	B
Cu74	1999	M	Hetero	Cuba	Cuba	1999	G
Cu85	1999	F	Hetero	Cuba	Cuba	1989	G
Cu87	1999	F	Hetero	Cuba	Cuba	1998	G
Cu103	1999	M	MSM	Cuba	Cuba	1996	BG
R77	1999	M	MSM	Spain	Cuba	na	BG
CB134	2003	M	MSM	Cuba	Cuba	2002	BG
CB378	2003	M	Hetero	Cuba	Cuba	2002	BG
CB471	2003	M	MSM	Cuba	Cuba	2003	BG
CB619	2003	M	MSM	Cuba	Cuba	2003	BG
CB118	2003	M	MSM	Cuba	Cuba	2000	BG
CB347	2003	M	MSM	Cuba	Cuba	2003	BG
CB418	2003	M	Hetero	Cuba	Cuba	2002	BG

*Subtype(s) in protease reverse transcriptase.

F indicates female; Hetero, heterosexual contact; M, male; na, not available.

bootstrap value (Fig. 2). Recombinants from the 3 subclades originally defined in *pol* again grouped within 3 distinct subclades (I, II, and III), each supported by a 100% bootstrap value. Subclades II and III were more related to each other than to subclade I. The Cuban BG recombinant clade clustered with the Cuban subtype G viruses. The 2 analyzed Cuban subtype B viruses formed a subclade within subtype B. For *gag-pol*, CB_418 clustered within subclade III with a 99% bootstrap value (not shown).

In the bootscan analyses (Fig. 3), Cuban viruses that were originally identified as nonrecombinant in partial *pol* sequences clustered within the same subtype all along the genome (see Figs. 3A, B). Consequently, they were used as the parental references in subsequent bootscan analyses to describe the mosaic structures of the Cuban recombinants. These analyses (see Figs. 3C–E) showed that (1) all BG recombinants were predominantly of subtype G, except for some B segments in the reverse transcriptase (polymerase and RNase H domains) and for subclade III recombinants also for a segment of *vif* of ~0.3 kb; (2) recombinants from the same subclade exhibited coincident mosaic structures; and (3) recombinants from different subclades had mosaic structures that were related but different.

In the bootscanning analyses, there were short segments in *pol* with an ambiguous subtype assignment (<70% bootstrap threshold; see Figs. 3C–E). Another short segment in *gp41* showed a decrease in bootstrap support for subtype G in all recombinants, accompanied by a short peak of support for subtype B. Suspecting that these might represent segments of recombinant origin but of shorter length than the specified (250 nucleotides) window size, new bootscanning analyses were performed with progressively decreasing window widths, which resulted in subtype assignments for these segments with bootstrap values ≥70% (not shown).

To map the recombination breakpoints more precisely, the programs Genconv,¹⁸ MaxChi,¹⁹ and GARD²⁰ were used. Breakpoint distribution determined by the various methods

was similar among viruses in the same subclade, although with minor differences resulting from sequence variation. Therefore, consensus breakpoints inferred by the 3 methods were obtained for each subclade, which were consistent with

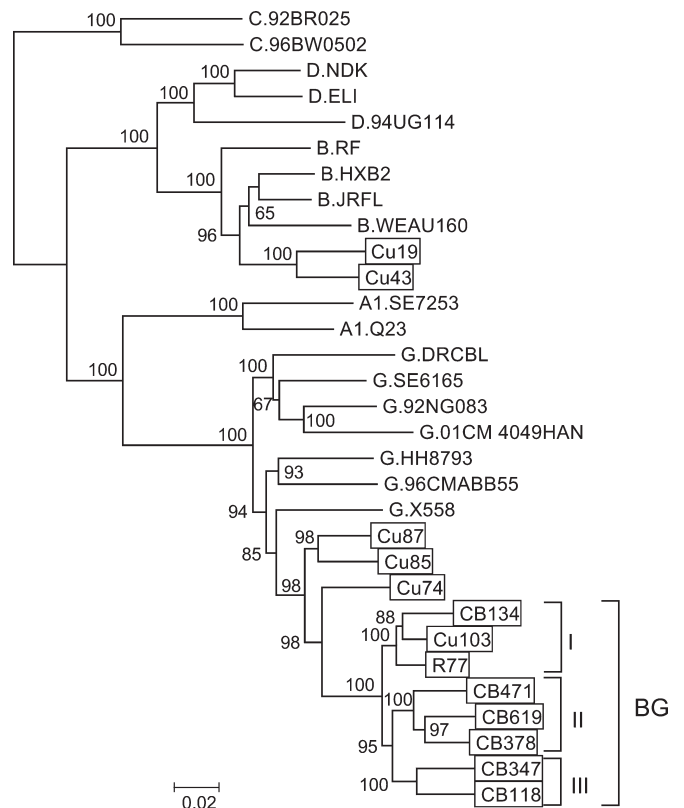


FIGURE 2. Maximum likelihood phylogenetic tree of near-complete genomes of BG recombinants and viruses of subtype B and G from Cuba. The names of Cuban viruses are boxed. Only bootstrap values ≥50% are shown.

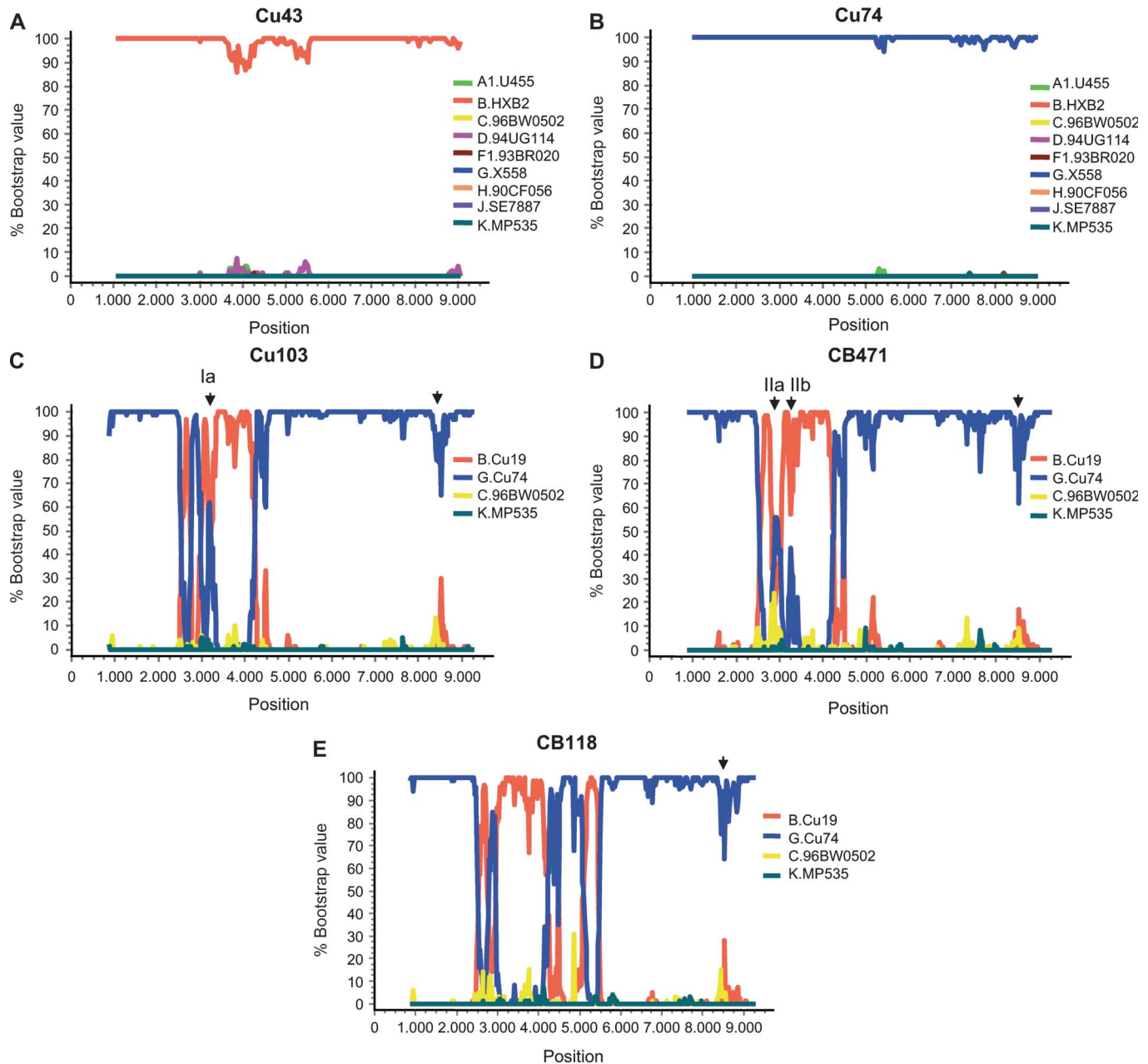


FIGURE 3. Bootscan analyses of near full-length genomes corresponding to a B subtype virus (Cu43) (A), a G subtype virus (Cu74) (B), and 3 putative BG recombinants from Cuba (Cu103, CB_471, and CB_118) (C–E), 1 from each of the 3 recombinant subclades. Windows of 800 nucleotides were used in A and B, and windows of 250 nucleotides were used in C through E, moving in 20-nucleotide increments. Trees were constructed with the neighbor-joining algorithm using the Kimura 2-parameter model of nucleotide substitution, with the transition/transversion ratio estimated from the data. Nucleotide positions in the horizontal axis coincide with the numeration in the HXB2 genome.

bootscanning analyses. There were 8 consensus breakpoints for each subclade: for subclade I, these were located in genome positions (the HXB2 isolate is used here and henceforth as a reference for numbering nucleotide positions in the HIV-1 genome) 2552, 2741, 2975, 3182, 3293, 4149, 8709, and 8758; for subclade II, these were located in genome positions 2552, 2843, 3014, 3293, 3380, 4158, 8697, and 8758; and for subclade III, these were located in genome positions 2552,

2745 through 2810, 2975 through 3029, 4149, 5151 through 5160, 5436, 8697, and 8751 through 8771 (the intervals correspond to sets of closely positioned breakpoints for which a unique precise consensus position could not be defined, presumably representing the same original breakpoint). Five of the consensus breakpoints were coincident among all recombinants from the 3 subclades and were located in nucleotide positions or intervals 2552, 2975 through 3029,

TABLE 2. SH Tests for Tree Topologies of Short (<200 Nucleotides) Fragments*

Fragments	Length	+G	+B
Subclade I			
2553–2741	189	0.042	0.517
3183–3293	111	0.508	0.024
8698–8750	53	0.016	1.000
Subclade II			
2844–3014	171	0.261	0.493
3294–3380	87	1.000	0.162
8698–8750	53	0.020	1.000
Subclade III			
2796–2975	180	1.000	0.084
8698–8750	53	0.018	0.768

*Implemented in PAUP*4.b10 using the MODELTEST best-fit model for the alignment. C.92BR025 and C.96BW050 were used as outgroups. The null hypothesis is represented by the best maximum likelihood tree with the constriction of clustering of the corresponding BG subclade with the subtype (B or G) indicated in the top row. The alternative hypothesis is represented by the unconstrained maximum likelihood tree. Heuristic searches for the maximum likelihood trees consisted of 1000 replicates of random sequence addition, with tree bisection and reconnection (TBR) branch swapping. The SH test 1-tailed *P* values were approximated with 1000 resampling of estimated log-likelihoods (RELL) bootstrap replicates. *P* values less than 0.1 are in bold type.

4149 through 4158, 8697 through 8709, and 8751 through 8771.

Subtype assignments for short segments (<200 nucleotides) were further analyzed by testing their phylogenetic congruence with the SH test (Table 2). Segments analyzed were 2553 through 2741 (Ia) and 3183 through 3293 (Ib) in

subclade I recombinants, 2844 through 3014 (IIa) and 3294 through 3380 (IIb) in subclade II recombinants, 2796 through 2975 (IIIa) in subclade III recombinants, and 8697 through 8748 in gp41 in all recombinants. Bootstrap supports $\geq 90\%$ were obtained in segment Ia for subtype B and in segments Ib, IIb, and IIIa for subtype G. Bootstrap support in segment IIa was only 60% for segment IIa and subtype G (Fig. 4). The SH tests rejected clustering with the alternative parental subtype, with significant values ($P < 0.05$) in segments Ia and Ib. The *P* value was marginally significant for segment IIIa ($P = 0.084$) and nonsignificant for segments IIa ($P = 0.261$) and IIb ($P = 0.162$).

In segment 8697 through 8748 in gp41, all BG recombinants clustered with Cuban subtype B with 92% bootstrap support (see Fig. 4F); alternative clustering with subtype G was rejected by the SH test in all 3 recombinant subclades ($P = 0.016$, $P = 0.018$, and $P = 0.020$, respectively). In addition, the subtype B origin for this short gp41 segment was supported by the presence of 3 highly unusual nucleotides found together only in the Cuban subtype B viruses: 8731G, 8732A, and 8747G (Fig. 5). Among all 4378 HIV-1 group M sequences comprising this segment retrieved from the Los Alamos HIV Sequence Database,⁴ each of these nucleotides was found in only 0.6%, 0.3%, and 0.6% of sequences, respectively, and no sequence contained more than 1 nucleotide. Thus, the combination of these 3 nucleotides represents a signature exclusive of the Cuban subtype B strain represented by Cu19 and Cu43 and the derived BG recombinants.

The analysis of signature nucleotides was also used to examine segments IIa and IIb, whose subtype affiliations were

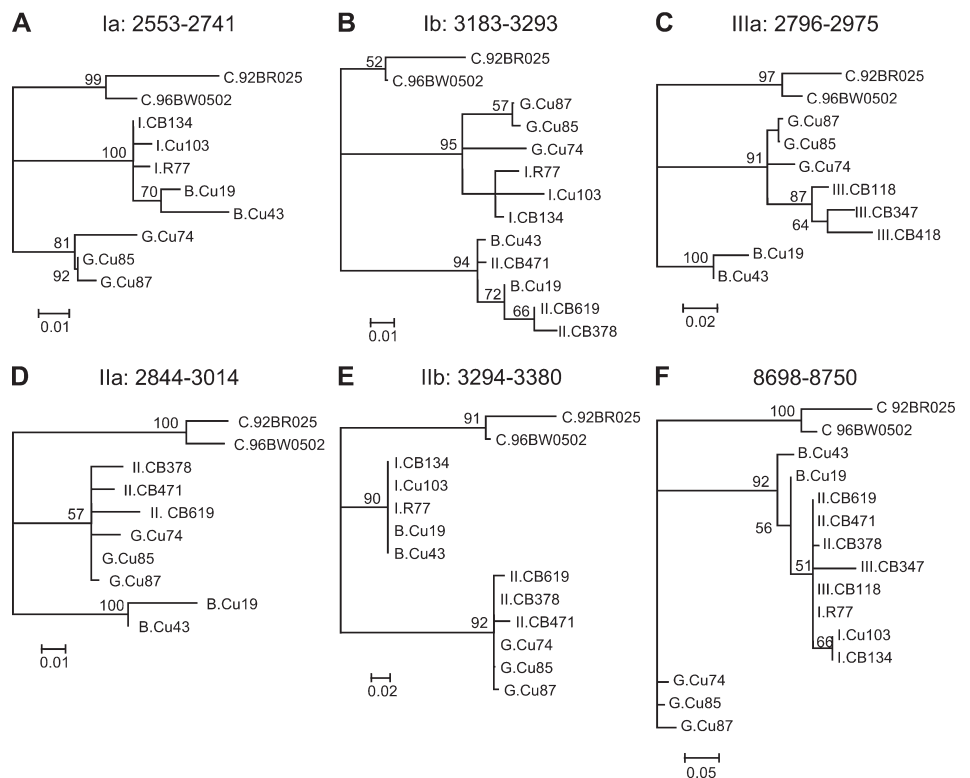


FIGURE 4. Maximum likelihood phylogenetic trees of short (<200 nucleotides) interbreakpoint segments from the BG recombinants. The fragment's designation as given in the main text, followed by positions in the HXB2 genome delimiting each segment, is shown above each tree. Only bootstrap values $\geq 50\%$ are shown.

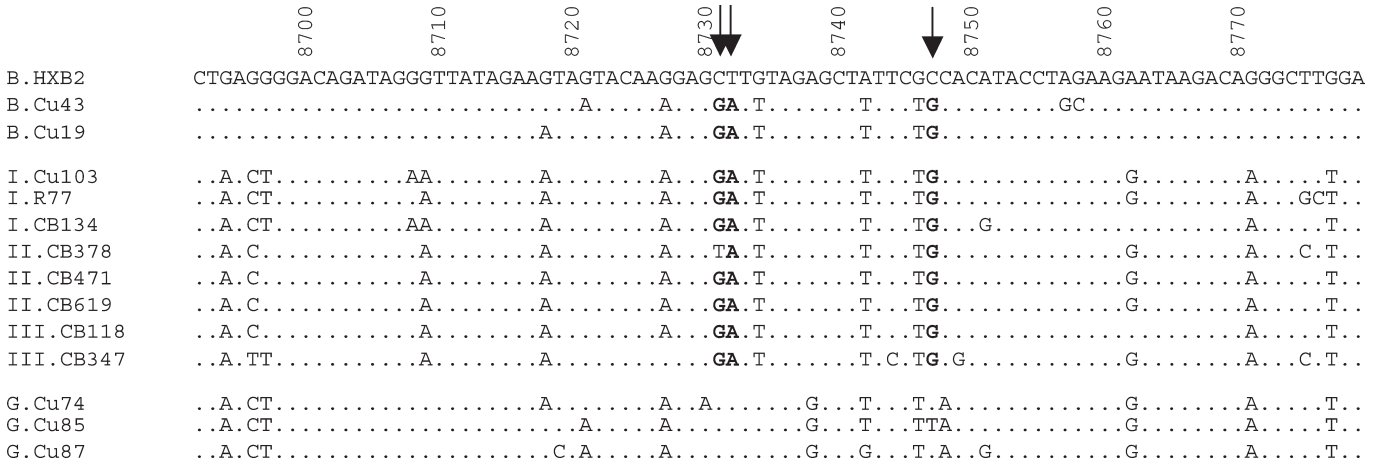


FIGURE 5. Alignment of the gp41 segment of the BG recombinants derived from subtype B. Dots indicate identity, with the HXB2 sequence on top. Signature nucleotides of the subtype B strain represented by Cu19 and Cu43 shared with the BG recombinants are in bold. Numbers above the alignment indicate positions in the HXB2 genome. Arrows signal the positions of the 3 signature nucleotides in the alignment.

unresolved by the SH tests. Segment IIa contained 5 nucleotide differences that were fixed between the Cuban subtype G and B parental strains. In subclade II, all 5 differences coincided with the subtype G parental strain (2876A, 2885G, 2921T, 2939C, and 2957T). When examining all subtype B sequences comprising this segment deposited in the Los Alamos Database, none of 23,536 sequences contained all 5 nucleotides (the maximum was 4, which were present in only 2 sequences). In segment IIb, there were 6 discriminative nucleotides, which, in subclade II, were fully coincident with the parental subtype G strain (3299A, 3317T, 3323A, 3339C, 3341A, and 3380A). Among 8190 subtype B sequences in the Los Alamos Database comprising segment IIb, none had more than 2 of these nucleotides. Therefore, the analysis of subtype

nucleotide signatures strongly suggests that segments IIa and IIb derive from the subtype G parental strain.

The mosaic structures of subclade I and II recombinants seemed to be similar, but there were 3 short segments in *pol* that showed phylogenetic incongruence: 2742 through 2843 (I-G, II-B), 3183 through 3293 (I-G, II-B), and 3294 through 3380 (I-B, II-G) (Fig. 6). In all 3 segments, SH tests significantly rejected tree topologies in which subclade I and II viruses cluster together. Importantly, for the near full-length genomes, subclade II is more related to subclade III than to subclade I (see Fig. 2). Taken together, these results indicate that despite substantial similarities in the recombinant structure, subclades I and II have distinct origins and represent different recombinant forms.

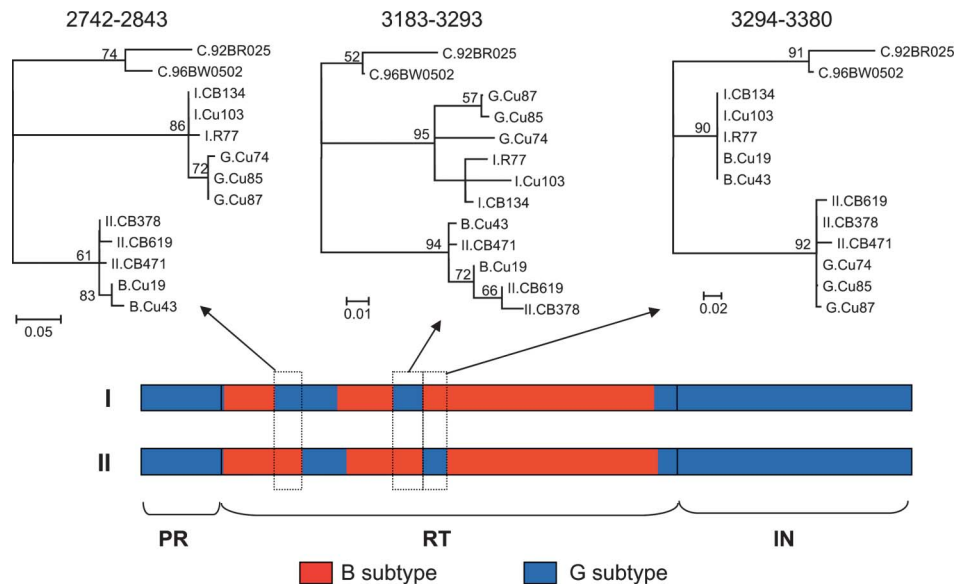


FIGURE 6. Maximum likelihood phylogenetic trees of 3 *pol* segments with discrepant topologies between subclades I and II. HXB2 positions delimiting the analyzed segments are shown on top. Only bootstrap values $\geq 50\%$ are shown.

To clarify further the phylogenetic relations within the Cuban BG recombinants and with the subtype B and G strains circulating in Cuba, we constructed 2 separate trees with the common concatenated subtype B and G segments, respectively (Fig. 7). In each tree, all Cuban BG recombinants formed a strongly supported clade. For the subtype B fragments (see Fig. 7A), subclade I is more closely related to subclade II (64% bootstrap), whereas for the subtype G fragments (see Fig. 7B), subclades II and III are closest (74% bootstrap). The trees also showed that subtype G segments of the recombinants are closely related to the Cuban subtype G variant represented by Cu74, Cu85, and Cu87 and that subtype B segments are closely related to Cuban subtype B strain represented by Cu19 and Cu43, which indicates that these 2 strains are the parental strains of the Cuban BG recombinants.

The tree of the subtype B segment in *vif*, which is present only in subclade III, shows that this segment also derives from the same parental strain that is ancestral to the subtype B portions of *pol* and *gp41* (Fig. 8).

The results shown here allow us to define 3 new CRFs, because 3 or more epidemiologically unlinked viruses with identical structures were characterized for each group of recombinants (at least 2 of each in >8 kb).³ According to the nomenclature system for HIV-1 CRFs, they were designated CRF20_BG, CRF23_BG, and CRF24_BG, corresponding, respectively, to subclades I, III, and II (numbered in the order in which sequences were obtained and analyzed). The mosaic structures of each CRF, inferred from the analyses reported in this article, are shown in Figure 9.

DISCUSSION

This study shows that the BG recombinants of Cuba analyzed here have diverse mosaic structures but that all derive

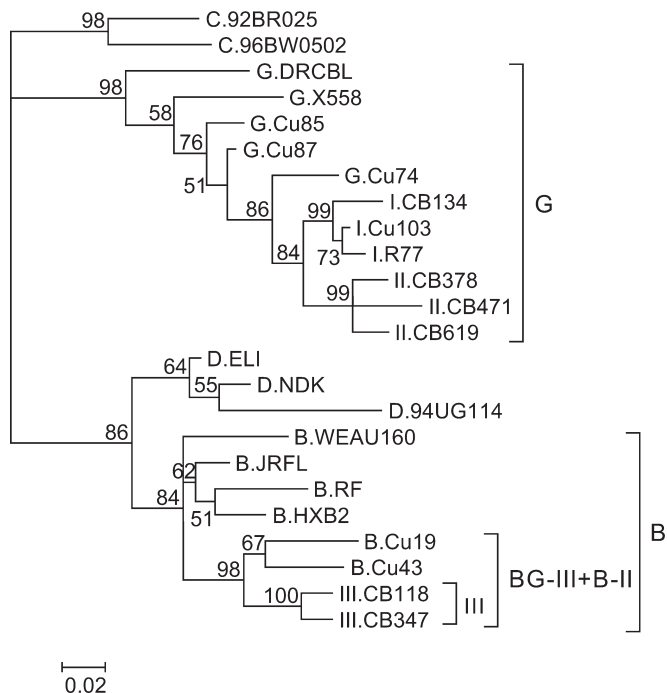


FIGURE 8. Maximum likelihood phylogenetic tree corresponding to the B subtype segment in *vif* for subclade III (positions 5151–5436). Only bootstrap values $\geq 50\%$ are shown.

from a common recombinant ancestor. This is inferred from substantial similarity of recombinant structures, with 5 coincident breakpoints, and from strongly supported clustering in phylogenetic trees of full-length genomes and of separate subtype G and B segments. Three different mosaic structures were identified in near full-length genomes, corresponding to the 3 subclades initially defined in *pol*

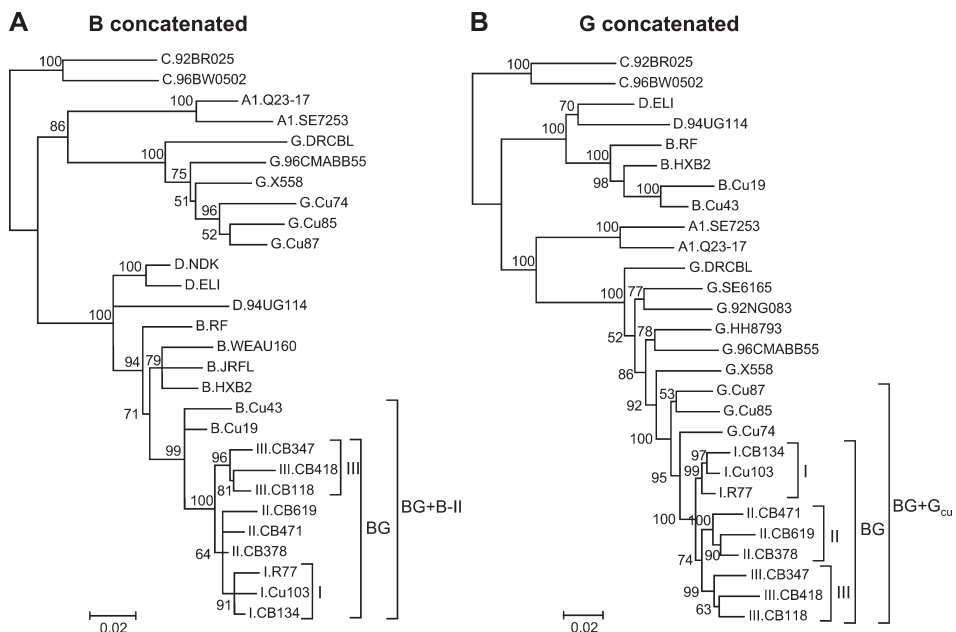


FIGURE 7. Maximum likelihood phylogenetic trees of concatenated B and G subtype segments of the Cuban BG recombinants. A, B subtype segments corresponding to HXB2 positions 2552 through 2741, 3014 through 3182, and 3380 through 4149. B, G subtype segments corresponding to positions 571 through 2551 and 4150 through 5096. Only bootstrap values $\geq 50\%$ are shown. B-II and G_{Cu} are the names given in Pérez et al⁶ to the Cuban strains represented by Cu19 and Cu43 in A and Cu74, Cu85, and Cu87 in B.

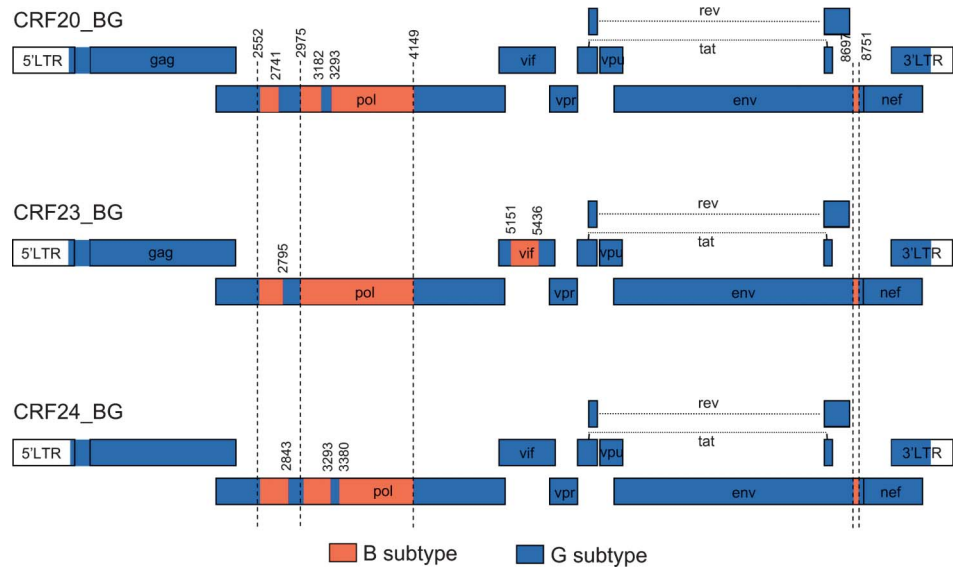


FIGURE 9. Schematic depiction of the mosaic structures of the 3 new CRFs identified in this study. Numbers above breakpoints refer to nucleotide positions in the HXB2 genome. Vertical dotted lines indicate breakpoints shared by all recombinants.

sequences, which allows us to define 3 new CRFs, because 3 epidemiologically unlinked viruses with identical structures were characterized for each subclade.

We also report the characterization of near full-length genomes of viruses representative of the subtype G variant circulating in Cuba. Phylogenetic analyses clearly indicate that this variant is one of the parental strains of the BG recombinants. These viruses form a distinct clade apart from other subtype G viruses (see Fig. 2), which supports their origin from a single introduction in Cuba. Viruses of this strain represented 4 (4%) of 105 analyzed in *pol* and V3 sequences from samples collected in 1999⁵ and 12 (2.8%) of 425 *pol* sequences from samples collected in 2003.⁶ In the most recent study,⁶ this strain was more prevalent among heterosexuals than among MSM (6.3% vs. 1.6%; $P < 0.05$, according to the Fisher exact test). It is the second subtype G variant circulating outside of Africa identified to date (after the variant circulating in Portugal and Galicia [northwestern Spain], from which CRF14_BG originated¹¹). Cuban and Iberian strains are closely related to each other (see Figs. 2, 7B). One possible explanation for this relationship is that the Iberian variant could derive from Angola, a former Portuguese colony, where a large number of Cubans were present during the 1980s.

We have also characterized near full-length genomes of subtype B viruses (Cu19 and Cu43) representative of the subtype B parental strain of the BG recombinants. In *pol*, Cu19 and Cu43 form a single clade with several other subtype B viruses from Cuba. Viruses of this strain (corresponding to the B-II cluster in a recent study⁶) represented 13 (7.4%) of 175 subtype B viruses in samples collected in 2003,⁶ and all were from MSM, 11 of whom were from Havana City.

Separate trees of B and G subtype segments suggest a closer relation of CRF20_BG and CRF24_BG in subtype B segments, whereas CRF23_BG and CRF24_BG are more closely related in subtype G segments (see Fig. 7). Discrepant topologies relating the recombinants in different segments of

the genome suggest a scenario of successive recombination events in which a common recombinant ancestor gave rise to diverse mosaic forms by secondary recombinations involving the original parental strains or other related recombinants. Three of the diverse recombinants generated in this process seem to have propagated to some epidemic extent, originating the 3 CRFs described here.

The origin and expansion of BG CRFs in Cuba seem to be recent. In samples collected in 2003, all individuals harboring BG recombinants were diagnosed with HIV-1 infection in 2000 or later, except for 2 in 1997 and 1 in 1996. Among MSM in Havana City, the population in which most of these recombinants were identified, there was a gradual increase of BG infections from 0% in those diagnosed in 1998 to 31.4% in those diagnosed in 2003,⁶ accompanied by a concomitant decline in the proportion of subtype B infections from 88.4% in 1997 diagnoses to 40% in 2003 diagnoses. The recent expansion of the BG recombinants among MSM in Havana City could derive from biologic features of these viruses resulting in increased transmissibility, from their chance introduction in promiscuous sexual networks practicing unprotected sex, or from both factors. The propagation of these recombinants in a relatively short period highlights the fact that the distribution of HIV-1 genetic forms in an area may change relatively rapidly over time, a phenomenon previously reported. Among these, the best known case is that of Thailand, where CRF01_AE, initially introduced among promiscuous heterosexuals, replaced subtype B as the predominant form transmitted among injecting drug users.²³ In Cuba, the expansion of the BG recombinant forms differs in that these have been generated locally *de novo*, probably among MSM, and have not propagated from other groups with different risk factors or other geographic areas. Rapid change in the distribution of genetic forms underscores the importance of maintaining continued surveillance to detect changes in the viral genetic composition of an epidemic promptly.

The BG recombinants of Cuba represent just another example of related HIV-1 recombinants, apparently derived from a common recombinant ancestor, circulating in an area, a phenomenon that has also been reported in other countries, such as Argentina,^{10,12} Brazil,²⁴ and southern China.^{25–27} This suggests a model of generation of these recombinants by successive rounds of recombination along diverse lineages,¹² which generates a great diversity of mosaic forms, only a few of which become circulating through the acquisition of novel biologic features or chance introductions into preexisting transmission networks (ie, founder events).

It is interesting to note that all but one of reported HIV-1 CRFs of non-African origin (03_AB, 07_BC, 08_BC, 12_BF, 14_BG, 15_01B, 28_BF, 29_BF, 31_BC, 33_01B, and the 3 Cuban BG recombinants identified here) have subtype B as the parental strain. This reflects the fact that subtype B is the HIV-1 genetic form with the most extensive geographic expansion and earliest propagation outside of Africa, which increases the chance of recombining with other genetic forms circulating in diverse areas.

With the identification of the 3 new CRFs and the characterization of Cuban subtype B and G strains, the number of HIV-1 genetic forms circulating in Cuba reported to date by characterization of near full-length genomes is 7: subtypes B and G, CRF18_cpx,⁷ CRF19_cpx,⁸ and 3 new BG CRFs. Subtype C is most likely also circulating in Cuba, although it has been characterized only in partial segments.⁶ Other genetic forms possibly circulating in Cuba but characterized only in partial segments are a B/CRF18 recombinant form and subtype H.⁶ The high genetic diversity of HIV-1 in Cuba is most remarkable, considering that the estimated number of prevalent HIV-1 infections in the country is only 4800,²⁸ and might be explained by the presence in the past of large numbers of Cubans in diverse sub-Saharan African countries and by the subsequent generation of recombinants among the introduced lineages.

In conclusion, we have characterized the mosaic structures and analyzed the phylogenetic relationship of HIV-1 BG intersubtype recombinant forms from Cuba, showing that these represent 3 new CRFs, all of which seem to derive from a common recombinant ancestor. We have also characterized the near full-length genome sequences of the Cuban parental subtype G and B strains, which has allowed us to define more precisely the location of the recombination breakpoints. Further work is needed to examine the associations of the newly identified genetic forms in Cuba with biologic properties, pathogenesis, transmission, immune responses, and response and resistance to antiretroviral drugs.

ACKNOWLEDGMENTS

The authors thank the United Nations Program on HIV/AIDS and, in particular, Saladin Osmanov and José Esparza for their support of this study. They also thank Pablo Martínez, Aurora de Miguel, and Alicia Ballester from the Genomic Unit at Centro Nacional de Microbiología, Instituto de Salud Carlos III, for technical assistance with sequencing.

REFERENCES

- Nájera R, Delgado E, Pérez-Álvarez L, et al. Genetic recombination and its role in the development of the HIV-1 pandemic. *AIDS*. 2002;16(Suppl 4):S3–S16.
- Thomson MM, Nájera R. Molecular epidemiology in the global AIDS pandemic: an update. *AIDS Rev*. 2005;7:210–224.
- Robertson DL, Anderson JP, Bradac JA, et al. HIV-1 nomenclature proposal. *Science*. 2000;288:55–56.
- HIV Sequence Database [database online]. Los Alamos, NM: Los Alamos National Laboratory. Available at: <http://www.hiv.lanl.gov/content/hiv-db/mainpage.html>. Accessed October 9, 2006.
- Cuevas MT, Ruibal I, Villahermosa ML, et al. High HIV-1 genetic diversity in Cuba. *AIDS*. 2002;16:1643–1653.
- Pérez L, Thomson MM, Bleda MJ, et al. HIV type 1 molecular epidemiology in Cuba: high genetic diversity, frequent mosaicism, and recent expansion of BG intersubtype recombinant forms. *AIDS Res Hum Retroviruses*. 2006;22:724–733.
- Thomson MM, Casado G, Posada D, et al. Identification of a novel complex circulating recombinant form (CRF18_cpx) of Central African origin in Cuba. *AIDS*. 2005;19:1155–1163.
- Casado G, Thomson MM, Sierra M, et al. Identification of a novel ADG intersubtype circulating recombinant form (CRF19_cpx) in Cuba. *J Acquir Immune Defic Syndr*. 2005;40:532–537.
- Boom R, Sol CJ, Salimans MM, et al. Rapid and simple method for purification of nucleic acids. *J Clin Microbiol*. 1990;28:495–503.
- Thomson MM, Delgado E, Herrero I, et al. Diversity of mosaic structures and common ancestry of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Argentina revealed by analysis of near full-length genome sequences. *J Gen Virol*. 2002;83:107–119.
- 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.
- Sierra M, Thomson MM, Rios M, et al. The analysis of near full-length genome sequences of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Chile, Venezuela and Spain reveals their relationship to diverse lineages of recombinant viruses related to CRF12_BF. *Infect Genet Evol*. 2005;5:209–217.
- Katoh K, Kuma K, Toh H, et al. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 2005;33:511–518.
- Jobb G, von Haeseler A, Strimmer K. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol*. 2004;4:18.
- Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics*. 1998;14:817–818.
- Salminen MO, Carr JK, Burke DS, et al. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses*. 1995;11:1423–1425.
- Lole KS, Bollinger RC, Paranjape RS, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol*. 1999;73:152–160.
- Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent recombination. *Virology*. 1998;265:41–48.
- Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci USA*. 2001;98:13757–13762.
- Kosakovsky Pond SL, Posada D, Gravenor MB, et al. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol*. 2006;23:1891–1901.
- Shimodaira H, Hasegawa M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol*. 1999;16:1114–1116.
- Swofford DL. *PAUP*: Phylogenetic Analysis Using Parsimony (and other methods)* [computer program]. Version 4b10. Sunderland, MA: Sinauer Associates; 2000.
- Subbarao S, Vanichseni S, Hu DJ, et al. Genetic characterization of incident HIV type 1 subtype E and B strains from a prospective cohort of injecting drug users in Bangkok, Thailand. *AIDS Res Hum Retroviruses*. 2000;16:699–707.

24. Sa Filho DJ, Sucupira MC, Casiero MM, et al. Identification of two HIV type 1 circulating recombinant forms in Brazil. *AIDS Res Hum Retroviruses*. 2006;22:1–13.
25. Yang R, Xia X, Kusagawa S, et al. On-going generation of multiple forms of HIV-1 intersubtype recombinants in the Yunnan Province of China. *AIDS*. 2002;16:1401–1407.
26. McClutchan FE, Carr JK, Murphy D, et al. Precise mapping of recombination breakpoints suggests a common parent of two BC recombinant HIV type 1 strains circulating in China. *AIDS Res Hum Retroviruses*. 2002;18:1135–1140.
27. Qiu Z, Xing H, Wei M, et al. Characterization of five nearly full-length genomes of early HIV type 1 strains in Ruili City: implications for the genesis of CRF07_BC and CRF08_BC circulating in China. *AIDS Res Hum Retroviruses*. 2005;21:1051–1056.
28. United Nations Program on HIV/AIDS. Available at: <http://www.unaids.org>. Accessed October 9, 2006.