

# Development and Application of a High-Throughput HIV Type 1 Genotyping Assay to Identify CRF02\_AG in West/West Central Africa

GUSTAVO H. KIJAK,<sup>1</sup> ERIC SANDERS-BUELL,<sup>1</sup> NATHAN D. WOLFE,<sup>2</sup> EITEL MPOUDI-NGOLE,<sup>3</sup> BOHYE KIM,<sup>1</sup> BRUCE BROWN,<sup>1</sup> MERLIN L. ROBB,<sup>4</sup> DEBORAH L. BIRX,<sup>4</sup> DONALD S. BURKE,<sup>2</sup> JEAN K. CARR,<sup>1</sup> and FRANCINE E. McCUTCHAN<sup>1</sup>

## ABSTRACT

In West/West Central Africa, CRF02\_AG is the most prevalent HIV-1 strain and circulates in the milieu of rare subtypes, circulating recombinant forms (CRFs), and unique recombinant forms (URFs). The molecular complexity of HIV-1 epidemics in this region and the need to extensively sample large populations, such as in the case of vaccine trials, pose seemingly conflicting requirements between full-genome sequencing and high-throughput low-resolution assays. Here we describe the development and evaluation of a multiregion hybridization assay (MHacr02) for the efficient genotyping of CRF02\_AG in West/West Central Africa. Subtype A, G, and CRF02\_AG-specific fluorescent probes were designed flanking five recombination breakpoints in CRF02\_AG and were used in real-time PCRs. A panel representing West/West Central African HIV-1 genetic diversity was evaluated by MHacr02. The sample set, previously characterized by full-genome sequencing, included CRF02\_AG and CRF02\_AG-containing recombinants ( $n = 28$ ), other subtypes, CRFs, and URFs ( $n = 34$ ). DNA from peripheral blood mononuclear cells, cocultures, and plasmids was used as template. When the patterns of probe reactivity were evaluated, CRF02\_AG was identified with a 100% specificity and sensitivity. In conclusion, MHacr02 will permit more efficient characterization of HIV-1 in West/West Central Africa, where CRF02\_AG is an important strain. Together with other regional genotyping assays MHacr02 will contribute to the development of a global picture of HIV-1 diversity and geographic distribution, providing a strong foundation for intervention, including vaccine development.

## INTRODUCTION

THE HIV-1 PANDEMIC is characterized by the circulation of multiple genetic subtypes and many more intersubtype recombinants.<sup>1-4</sup> Some of the recombinants have been found in only one patient [i.e., unique recombinant forms (URFs)], while others have widely spread in populations [i.e., circulating recombinant forms (CRFs)].

Extensive HIV-1 genetic diversity is represented in the strains that have been found in West/West Central Africa.<sup>5-25</sup> In this region, CRF02\_AG, a recombinant with a complex mosaic genomic structure, is the predominant strain, circulating in

the milieu of many other subtypes, CRFs, and URFs. Based on previous reports, the proportion of HIV-1 infections with CRF02\_AG ranges from 95 to 50% in different West/West Central African countries.<sup>5-25</sup> Although the surveillance has not been systematic, and various genotyping approaches have been used with variable power to discriminate the CRF02\_AG from other forms, it appears that CRF02\_AG reaches its highest proportion of strains in the far West, with a gradual decrease to 50% or less in Central Africa.<sup>5-27</sup> Another important finding has been that no single other clade accounted for the rest of the HIV-1 cases.<sup>6-8,11,12,19,23-25</sup> CRF02\_AG is typically embedded in epidemics with a very complex mixture of rare subtypes and

<sup>1</sup>The Henry M. Jackson Foundation, Rockville, Maryland 20850.

<sup>2</sup>Johns Hopkins School of Public Health, Baltimore, Maryland.

<sup>3</sup>Hopital Militaire de Yaounde, Yaounde, Cameroon.

<sup>4</sup>Walter Reed Army Institute of Research, Rockville, Maryland.

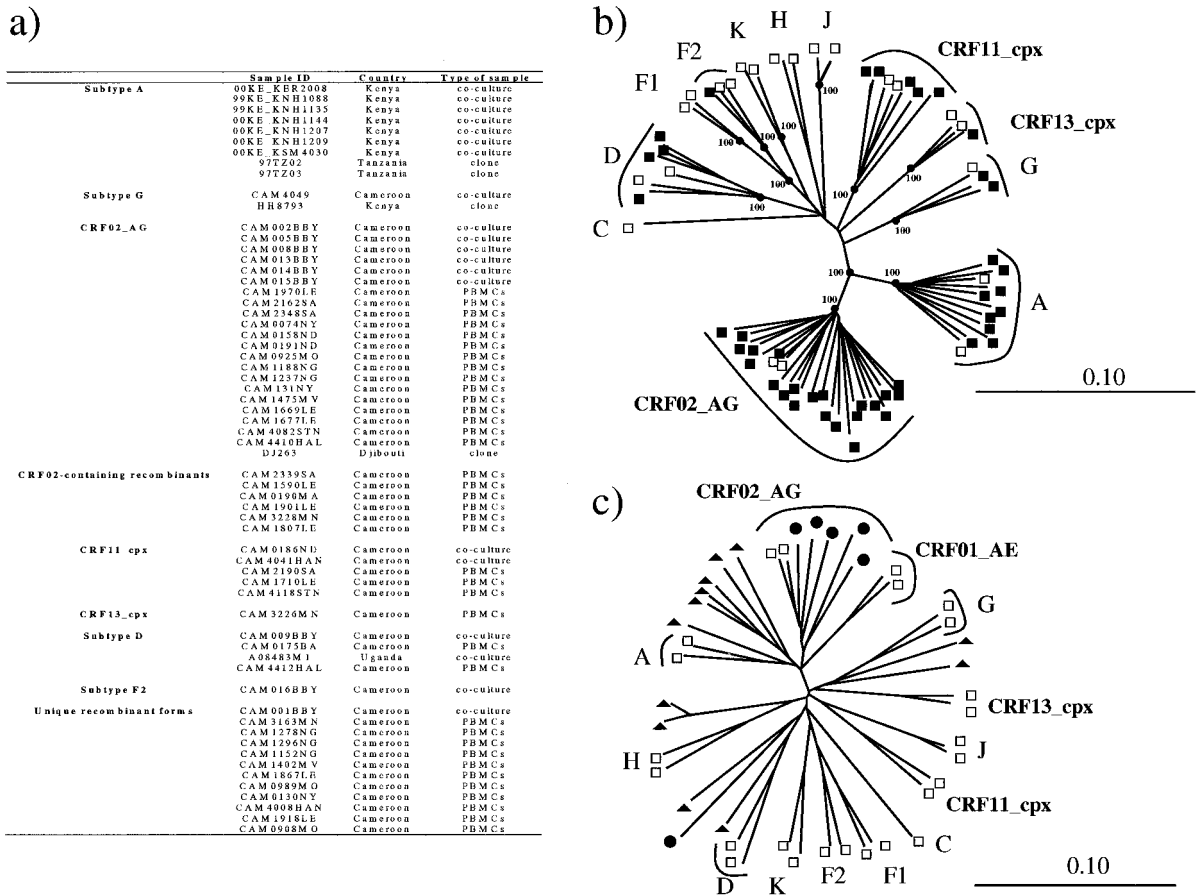
other recombinants. This has also led to the emergence of “second-generation” recombinants between the CRF and the other clades.<sup>7,28,29</sup>

CRF02\_AG is one of six globally prevalent HIV-1 strains<sup>1–4</sup> and, accordingly, it has been selected for use in candidate HIV-1 vaccines. In the context of cohort development for interventions, there is a great need for a high-throughput genotyping tool that can clearly describe the molecular picture in the region. There seem to be conflicting methodological requirements for such an assay. On the one hand, it should be able to distinguish between CRF02\_AG and related strains, and all other clades circulating in the region (including A/G recombinants not related to CRF02\_AG). Currently, this can be achieved only by full-genome sequencing, a low-throughput and expensive technique.<sup>30–32</sup> On the other hand, genotyping of large cohorts can be assessed only by high-throughput tests, which have the drawback of low genetic resolution. Partial sequencing and het-

eroduplex mobility assay (HMA) are among the latter genotyping tools that have been applied in the study of CRF02\_AG epidemics.<sup>5,6,8,10,12,25,33–36</sup>

A novel group of regional genotyping tests, the multiregion hybridization assays (MHAs), seems to bridge this methodological gap.<sup>37</sup> These assays are based on the hybridization of clade-specific fluorescent probes in a real-time PCR. MHAs are being used to describe complex epidemics, each tailored to the limited genetic diversity in the regions: subtypes A, C, and D in East Africa,<sup>37</sup> B and F in South America,<sup>38</sup> and B, C, and CRF01\_AE in Southeast Asia,<sup>39</sup> respectively.

Utilizing a similar MHA approach, we present the development of MHAcfr02, a high-throughput genotyping tool for the study of CRF02\_AG in the West/West Central African HIV-1 epidemics. The simultaneous assessment of different HIV-1 genomic regions by a real-time PCR-based technique confers a high sensitivity and specificity to the assay, which make it suit-



**FIG. 1.** Samples employed in the development of MHAcfr02. (a) List of the genotypes of the specimens, their country of origin, and type of sample. The CRF02\_AG genomic content was 80% for CAM2339SA, 75% for CAM1590LE, 65% for CAM0190MA, 55% for CAM1901LE, 90% for CAM3228MN, and 40% for CAM1807LE. Unique recombinants contained subtypes A, D, F2, G, and CRF01\_AE. Accession numbers of newly and previously reported sequences are listed in Materials and Methods. (b) Neighbor-joining tree based on full-length sequences of subtype A, G, D, F2, CRF02\_AG, CRF11\_cpx, and CRF13\_cpx strains. The tree was built using Kimura two-parameter distance method (transition/transversion = 2).<sup>45</sup> Strains studied by MHAcfr02 are shown as (■) and reference sequences are shown as (□). Bootstrap values supporting clade-defining nodes are shown (over 100 replicates). (c) Neighbor-joining tree of CRF02\_AG-containing recombinants and unique recombinant forms, represented as (●) and (▲), respectively.

able for the support of population-based interventions in the region, including the development of HIV-1 vaccines.

## MATERIALS AND METHODS

### Samples

In the development of the MHAcrrf02, a panel of 62 samples, from which a virtually full-length sequence had been obtained, was used. Samples were chosen so that they represent the HIV-1 genetic diversity in West/West Central Africa, and included CRF02\_AG, subtypes A, D, F2, and G, CRF11\_cpx, CRF13\_cpx, and URFs (Fig. 1a). CRF02\_AG-containing second-generation recombinants were also represented in the panel. Samples were collected mainly in Cameroon, and specimens from Kenya, Tanzania, Uganda, and Djibouti were also used. Depending on availability, primary peripheral blood mononuclear cells (PBMCs), cocultures, or nearly full-length HIV-1 genomes cloned in plasmids (TOPO-TA vectors, Invitrogen, Carlsbad, CA), were employed.

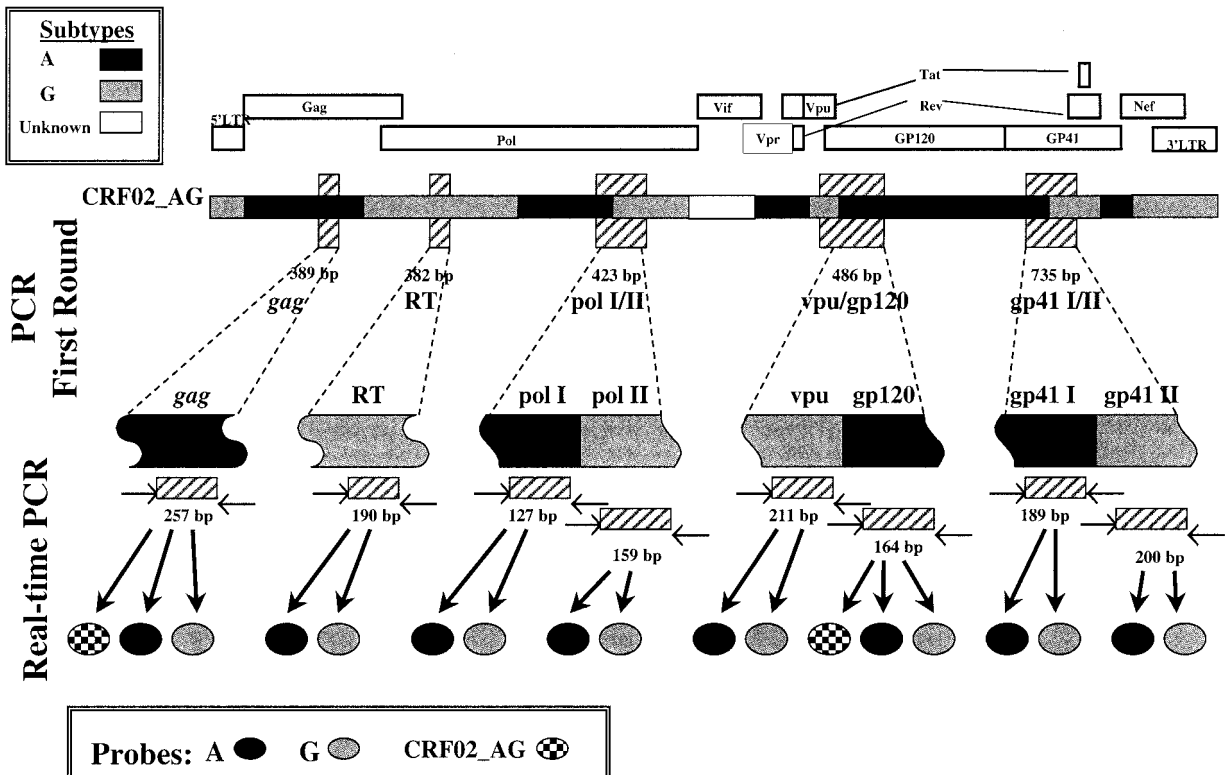
PBMC DNA was extracted using the MagnaPure robotic nucleic acid extraction procedure (Roche Diagnostics Corp., Indianapolis, IN), following the manufacturer's instructions. Coculture and plasmid DNA were extracted manually using

QIAamp DNA and QIAGEN Plasmid kits (Qiagen Inc., Valencia, CA), respectively.

Nearly full-length HIV-1 genome sequencing was performed as previously described.<sup>30,31</sup> Sequences were manually aligned with reference strains of relevant subtypes and CRFs.<sup>40</sup> Phylogenetic relatedness was established using DNADIST, NEIGHBOR, SEQBOOT, and CONSENSE, included in the PHYLIP software package version 3.2c.<sup>41</sup> The obtained trees were visualized using TREE TOOL.<sup>42</sup> Bootscanning (sliding window of 300 bases) was used to confirm the genomic structures.<sup>43</sup>

### MHAcrrf02

The principle of the MHAcrrf02 is to amplify short fragments along the HIV-1 genome and assess the hybridization of clade-specific fluorescent probes in a real-time PCR.<sup>37</sup> The amplified regions were positioned flanking recombination breakpoints that have been previously mapped on the CRF02\_AG genomic structure.<sup>7,18</sup> The assay is performed in a nested format. In separate first-round PCRs, five regions of the HIV-1 genome (the *p24/p2/p7* part of *gag*, the 5' part of the RT, the RnaseH/INT part of *pol*, the *vpu/5'* extreme of *gp120*, and part of *gp41*) are amplified using universal primers (Fig. 2). Then, each amplicon is distributed in one (in the cases of the *gag* and the RT regions) or two (in the cases of RnaseH/INT part of *pol*,



**FIG. 2.** Design of the multiregion hybridization assay for CRF\_02 (MHAcrrf02). Five regions of the HIV-1 genome flanking recombination breakpoints in CRF02\_AG are amplified in a first-round PCR. Real-time PCRs with clade-specific fluorescent probes that hybridize at either side of the breakpoints are performed. Probes specific for subtypes A, G, and CRF02\_AG are used in separate reactions. The names of the different amplicons generated in the PCRs and their sizes are given. The mosaic genomic structure of CRF02\_AG and the mapping of HIV-1 genes are shown on the top. See text for details.

*vpu/gp120*, and *gp41*) separate sets of second-round PCRs. Each of these reactions is performed in a Taqman real-time format, with universal primers, and fluorescent probes with specificity for subtypes A or G. When possible, CRF02\_AG-specific probes are also used, taking advantage of nucleotide substitutions that distinguish the subtype A portion of CRF02\_AG. Probes for each clade are used in separate real-time PCR reactions. All probes were labeled at the 5' end with 6-carboxyfluorescein (FAM) and had Black Hole Quencher (BHQ) at the 3' end (Biosource, Camarillo, CA). Probes were designed to have melting temperatures ( $T_m$ ) between 61 and 66°C. Probe sequences were selected considering regions of high within-clade conservation but that contained at least three mismatches between subtypes distributed along the probe sequence. When necessary, to improve the probe specificity, a deliberate mismatch was introduced so that the probe would still hybridize to homotypic strains but not to heterotypic ones, as suggested by de Baar *et al.*<sup>44</sup> The sequences of primers and probes used in MHAcrf02 are shown on Table 1.

The first-round PCR contained GeneAmp 10X PCR Buffer (Applied Biosystems, Foster City, CA), 250  $\mu$ M each dNTP, 1.5 mM MgCl<sub>2</sub>, 400 nM of each outer primer, 2.5 U of AmpliTaq DNA Polymerase (Applied Biosystems), and 10  $\mu$ l of extracted DNA, in a final volume of 50  $\mu$ l. The thermocycle routine was 1 cycle at 95°C for 10 min, then 30 cycles of 95°C 10 sec, 55°C 30 sec, and 72°C 1 min, and a final extension at 72°C for 7 min, in an ABI 9600 thermocycler. The second-round PCRs contained Taqman 2x Universal PCR Master Mix (Applied Biosystems), 400 nM of each inner primer, 250 nM probe, and 5  $\mu$ l of first-round PCR product, in a final volume of 25  $\mu$ l. PCR amplification was performed in a 96-well ABI PRISM 7900HT Sequence Detection System (Applied Biosystem) at 95°C 15 sec, 55°C 30 sec, and 60°C 1 min for 40 cycles. Fluorescence intensity was monitored during the reaction and was analyzed using the SDS v2.1 software (Applied Biosystem). The threshold was set based on negative controls and results were considered positive when the threshold cycle ( $C_T$ ) was lower than 30 and if an exponential increase in the normalized fluorescence intensity for over five consecutive cycles was observed. Negative real-time PCR results due to lack of sample amplification were assessed by electrophoresis in a 2% agarose gel.

The lower detection limit of MHAcrf02 as determined by the reactivity of serial dilutions of HIV-1 plasmids was 7–70 input copies (data not shown).

## RESULTS

### *Evaluated samples*

The aim of this study is to develop a high-throughput assay that can distinguish CRF02\_AG from other subtypes, CRFs, and URFs that circulate in West/West Central Africa. For this purpose, we assembled a panel of samples from which a virtually full-length HIV-1 sequence had been obtained and that included much of the HIV-1 diversity in the region.

In Figure 1, 62 strains employed in the development of the assay are shown. The panel included CRF02\_AG ( $n = 22$ ), the parental subtypes A ( $n = 9$ ) and G ( $n = 2$ ), as well as other

clades frequently found in West/West Central Africa, such as CRF11\_cpx ( $n = 5$ ), CRF13\_cpx ( $n = 1$ ), subtype D ( $n = 4$ ), F2 ( $n = 1$ ), and unique recombinants between subtypes A, D, F2, G, and CRF01\_AE ( $n = 12$ ). Accounting for second-generation recombinants between CRF02\_AG and other subtypes that have been described,<sup>7,28,29</sup> six samples that contained 40–90% of CRF02\_AG in their genomic structure were also included. The type of sample used and their countries of origin are shown in Figure 1a. The phylogenetic identity of the strains analyzed is shown on Figure 1b and c.

### *Design of the assay*

Five regions along the HIV-1 genome that included the HIV-1 structural genes *gag*, *pol*, and *env* were selected based on the recombinant structure of CRF02\_AG (Fig. 2). Universal first-round PCR primers were located flanking known recombination breakpoints in CRF02\_AG. Each amplicon was then distributed into several second-round PCRs, each with a clade-specific fluorescent probe, in a Taqman real-time PCR format.

In the first-round PCR, five different amplicons were generated: *gag*, RT, *pol* I/II, *vpu/gp120*, and *gp41* I/II. In the cases of *pol* I/II, *vpu/gp120*, and *gp41* I/II, each amplicon was then distributed into two separate sets of real-time PCRs (*pol* I and *pol* II, *vpu* and *gp120*, and *gp41* I and *gp41* II, respectively), each with its own set of inner primers. In that way the genotype at each side of the recombination breakpoint could be assessed. The *gag* and RT amplicons were each distributed into a single set of corresponding real-time PCRs (*gag* and RT, respectively). Thus, eight different sets of second-round real-time PCRs were designed.

In the second-round PCRs, 16 fluorescent probes that had specificity for clades A or G were used. In addition to that, two CRF02\_AG-specific probes were designed based on the differences that exist within the clade A radiation between the CRF02\_AG lineage and other lineages (i.e., CRF01\_AE and East African subtype A).<sup>7,18</sup> As a consequence of this, 18 probes were designed. Each probe was used in a separate PCR.

In all the cases, inner and outer primers were placed in areas of high sequence conservation between subtypes, whereas the probes were located in regions of high within-clade conservation but high between-subtype divergence. The inclusion of mixed bases was kept to a minimum and limited to the cases when it was necessary to increase the universality of the primers or the sensitivity of the probes. The nucleotide sequences of outer and inner primers and probes are shown in Table 1.

### *Performance of MHAcrf02 on the tested panel*

The performance of the MHAcrf02 was assessed on a panel of 62 samples. The overall rate of reactivity for the eight regions tested was 467/496 (94.15%). When each of the amplified regions was assessed separately, the PCR reactivity was 87, 97, 97, 89, 95, 97, 91, and 89% for *gag*, RT, *pol* I, *pol* II, *vpu*, *gp120*, *gp41* I, and *gp41* II, respectively. No statistically significant difference was found between the PCR reactivity for CRF02\_AG samples when compared to the other subtypes and recombinant forms. In pattern recognition-based assays such as MHAcrf02, the number of reactive PCRs per sample is also important. In the current assay, 54/62 samples (87%) had at least

TABLE 1. PRIMERS AND PROBES USED IN THE MULTIREGION HYBRIDIZATION ASSAY FOR CRF02\_AG (MHAcrf02)

	Name	Sequence (5' → 3') <sup>a</sup>	Forward/reverse
<i>gag</i>			
Outer primers	GAGOF	GATGACAGAYACCTTRYTGTTCCA	Forward
	GAGOR	CCTGGCCTCCCYTTGYTGAA	Reverse
Inner primers	GAGIF	GGGCTACATTAGAAGAAATGATGACAGCA	Forward
	GAGIR	CARTCTTCATTTGRTGTCCYTCYTT	Reverse
Probes			
CRF02	GAG-02	CARTTGAACAYTTTATTGTTCTCTGGC	Reverse
A	GAG-A	CARTTGAACAYTTWATYSTTYTCTGGC	Reverse
G	GAG-G	CAATTGCAACAGTTTRATADTTCTTYTYGGGC	Reverse
RT			
Outer primers	RTOF	GCCATTGACAGAAGARAAAATAAAAGC	Forward
	RTOR	CCTGTGGAAGCACATTGTAYTGATA	Reverse
Inner primers	RTIF	GGGCCTGAAAATCCATAYAATACTC	Forward
	RTOF	GCRTCCCYACATCTAGTACTGTT	Reverse
Probes			
A	RT-A	TTARRCCTGCTGGATGYGGTATTC	Reverse
G	RT-G	TTAAYCCYGCRGTATGAGGTATYC	Reverse
<i>pol III</i>			
Outer primers	POLOF	ACAAAYCAAAAGACTGAATTACATGCAAT	Forward
	POLOR	TGCCATATYCCCTGGACYACAGTCTAC	Reverse
<i>pol I</i>			
Inner primers	POLI-IF	GCACAACCAGACAGGAGTGAATCAGA	Forward
	POLI-IR	TTATCTACTGTTCATTTCTCCAATTCC	Reverse
Probes			
A	POLI-A	AGGTAGACTTTGTCCTTTYCTATTARCTTCTC	Reverse
G	POLI-G	AGGTAGACCTTTTCCTTTTTATTAGCTGTTC	Reverse
<i>pol II</i>			
Inner primers	POLII-IF	GTACCAGCACACAAAGGGATTGG	Forward
	POLII-IR	GCAGATTAATACTACTAGCCATTGCTCT	Reverse
Probes			
A	POLII-A	CATGTTCTTAATTGAGCTTTATCTATCCCATC	Reverse
G	POLII-G	CATGVTCTTAATTGGCTTTATAATATGCCATC	Reverse
<i>vpu/gp120</i>			
Outer primers	VPGPOF	GCATYTCTATGGCAGGAAGAAGC	Forward
	VPGPOR	GCATGTGTRGCCCARACATTATG	Reverse
<i>vpu</i>			
Inner primers	VPUIF	CATCAAAATCCTRTACCAAAGCAGTRAGT	Forward
	VPUIR	TCACTCTCATTNCCACTRTCTTCTGC	Reverse
Probes			
A	VPU-A	CCACACAACTATTGCTAKGATTAGYGC	Reverse
G	VPU-G	CACAAAYTRTGCTGCTATGAATGCTAC	Reverse
<i>gp120</i>			
Inner primers	GP120IF	CAGAAGAYAGTGGNAATGAGAGTGA	Forward
	GP120IR	GGTNTCTGCRTCTYYCCACACAGGTA	Reverse
Probes			
CRF02	GP120-02	YGTGACCCACAANTYKTYAGCATTACA	Reverse
A	GP120-A	ACCCACAAKTKTCTGCAGYACTACA	Reverse
G	GP120-G	TRACCCACAARTTATYTAAGGCACTACA	Reverse
<i>gp41 I/II</i>			
Outer primers	GKGP410F	GAGCAGCAGGAAGCACTATGG	Forward
	GKGP410R	TGGTAGCTGAAGAGGCACAGG	Reverse
<i>gp41 I</i>			
Inner primers	GP41I-IF	GTCAATAACGCTGACGGTACAGG	Forward
	GP41I-IR	CCAAATYCCTAGGAGCTGTTGATC	Reverse
Probes			
A	GP41I-A	CAGATGTTGTTGAGCCTCAATAGCCYTCA	Reverse
G	GP41I-G	CAGATGCTGCTGCGCCTTATAG	Reverse

(continued)

TABLE 1. PRIMERS AND PROBES USED IN THE MULTIREGION HYBRIDIZATION ASSAY FOR CRF02\_AG (MHAcrf02) (CONT'D)

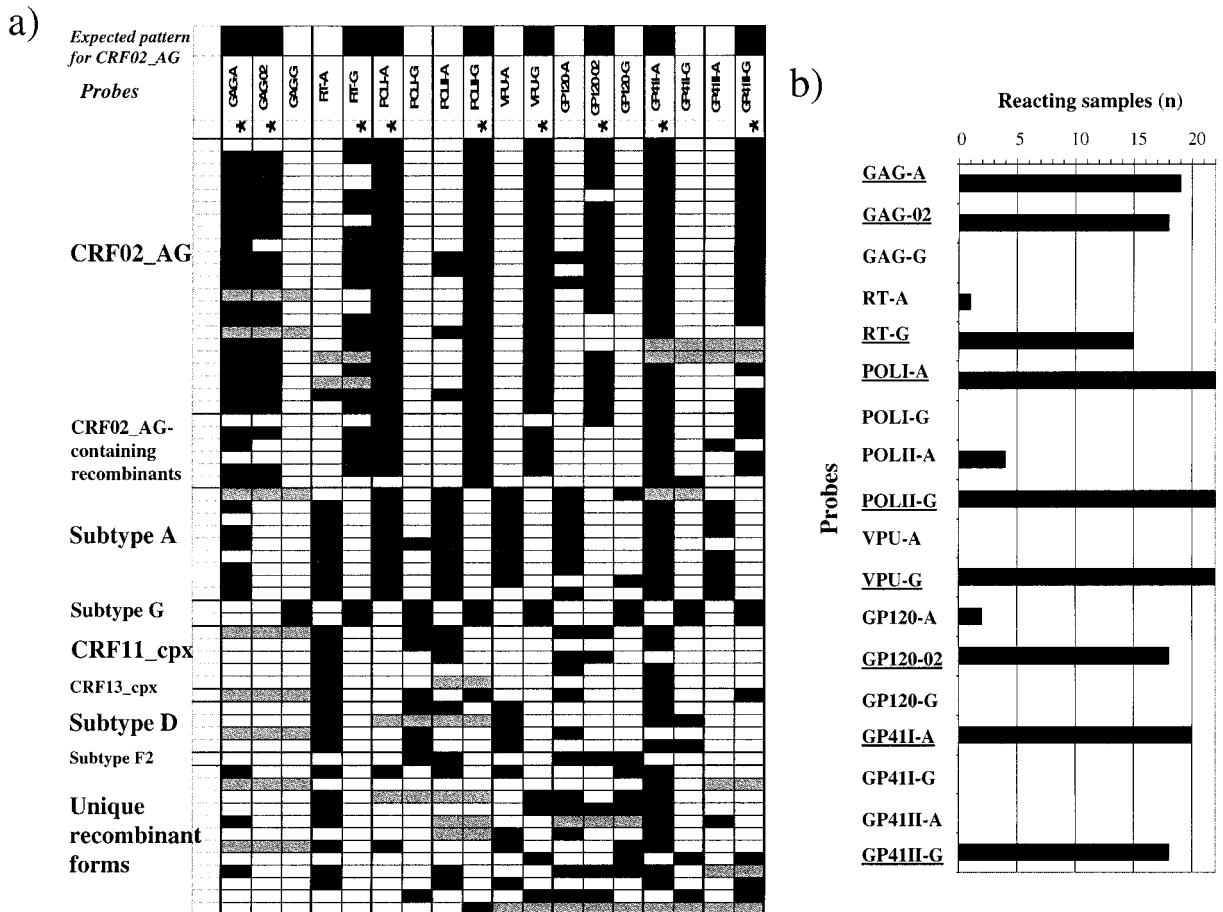
	Name	Sequence (5' → 3') <sup>a</sup>	Forward/reverse
<i>gp41</i> II			
Inner primers	GP41II-IF GP41II-IR	CAARNTGGCTRTGGTATATAARAATAT GTYTGYCTTGCTCKCCACCT	Forward Reverse
Probes			
A	GP41II-A	CAGAYCCATNYCCAAACCCARG	Forward
G	GP41II-G	CAGAYCCTTACCCACCACCAGMGG	Forward

<sup>a</sup>Deliberately induced mismatches are underlined.

seven positive PCRs out of eight, with no significant differences among clades. These results clearly indicate the sensitivity of MHAcrf02.

The observed probe-reactivity patterns for the studied panel are shown on Figure 3a. The 22 CRF02\_AG strains shared a pattern consistent with the structure of CRF02\_AG: they reacted with probes GAG-02, GAG-A, RT-G, POLI-A, POLII-G, VPU-G, GP120-02, GP41I-A, and GP41II-G (which will be

referred to as “homologous” probes), and they did not react with probes GAG-G, RT-A, POLI-G, POLII-A, VPU-A, GP120-A, GP120-G, GP41I-G, and GP41II-A (which will be referred to as “heterologous” probes). Within the CRF02\_AG cluster, minor deviations from this pattern were observed and were caused by failure in PCR, lack of “homologous”-probe reactivity, or unexpected reaction with “heterologous” probes, as previously reported for other MHA systems.<sup>37</sup> In the case of “second-gen-



**FIG. 3.** Performance of MHAcrf02 on the assessed panel ( $n = 62$ ). (a) Probe reactivity patterns of samples sorted by genotype. Black, white, and gray boxes represent positive probe reaction, lack of probe reaction, and lack of PCR reaction, respectively. The expected probe-reactivity pattern for CRF02\_AG is shown on the top, and homologous probes are depicted with an asterisk. (b) Analysis of the reactivity of MHAcrf02 probes on CRF02\_AG samples ( $n = 22$ ). Homologous probes are underlined. See text for details.

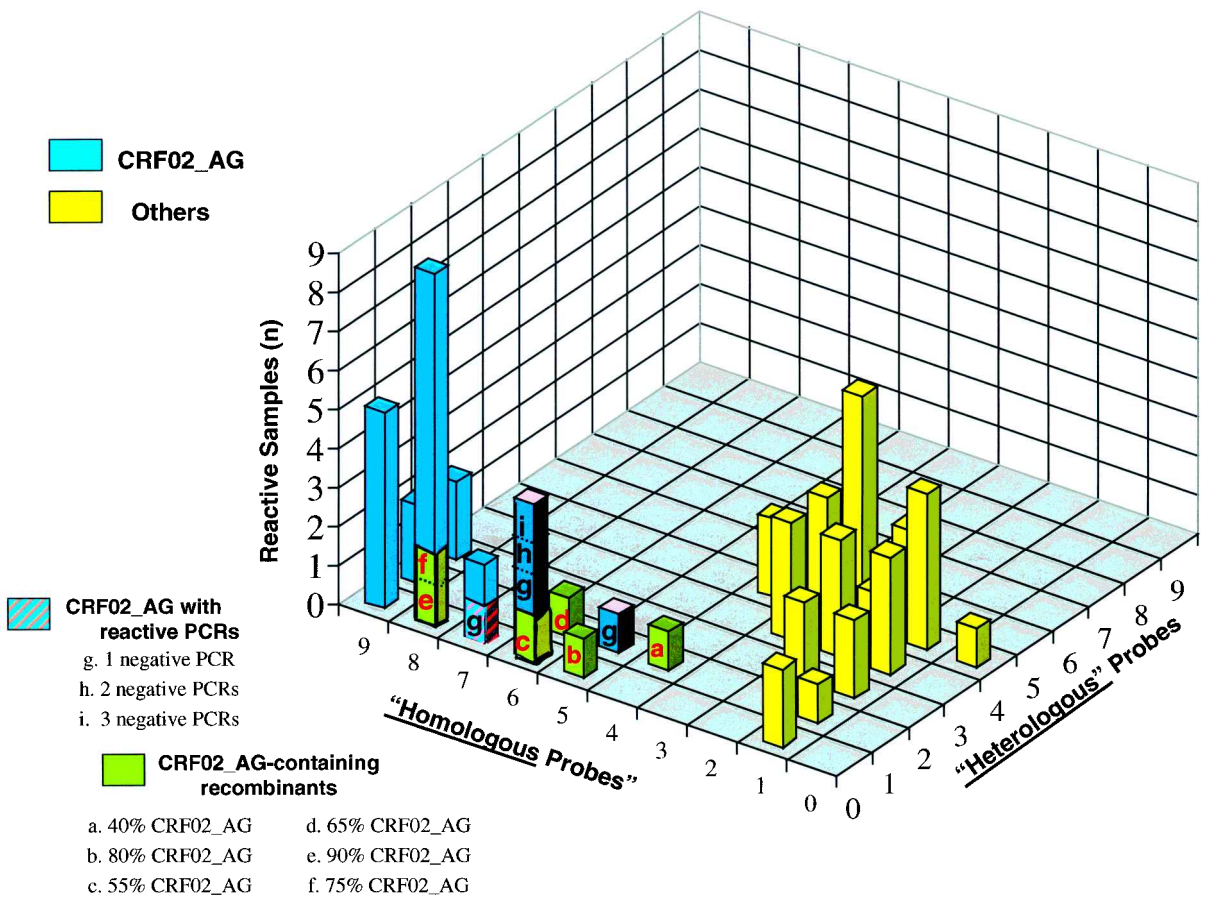
eration” recombinants, parts of the CRF02\_AG pattern were conserved, which allow for the characterization of these strains as related to CRF02\_AG.

The CRF02\_AG probe-reactivity pattern was specific for this clade and different from that of parental subtypes A and G, and other relevant subtypes, CRFs and URFs (Fig. 3a). Note that a level of reactivity for non-CRF02\_AG samples with “homologous” probes is expected, for example, subtype A samples are expected to react with three “homologous” probes (i.e., GAG-A, POLI-A, and GP41I-A) and five “heterologous” probes (i.e., RT-A, POLII-A, VPU-A, GP120-A, and GP41II-A). Other subtypes, CRFs, and URFs may react with a number of “homologous” and “heterologous” probes depending on the A- or G-subtype genomic content and the homology with MHAcrrf02 probes.

The analysis of the specificity of the MHAcrrf02 probe system for CRF02\_AG is shown on Figure 3b. For 22 CRF02\_AG samples, the average reactivity with “homologous” probes was 88% (range: 68–100%), while the average reactivity with “heterologous” probes was 4% (range: 0–18%).

*Establishment of criteria for identification of CRF02\_AG by MHAcrrf02*

To clearly establish a criterion for classifying a sample as CRF02\_AG or a CRF02\_AG-containing recombinant, an analysis of the number of “homologous” and “heterologous” probes reacting in samples of different clades was carried out. In Figure 4, two separate groups were defined. On the one hand, CRF02\_AG samples and CRF02\_AG-containing recombinants grouped together in a region defined by four to nine reactive “homologous” probes and zero to two “heterologous” probes. Within this group, three subclusters were defined based on their PCR and probe reactivity in MHAcrrf02 and their genomic structure: CRF02\_AG with eight reacting PCRs ( $n = 17$ ), CRF02\_AG with five to seven reacting PCRs ( $n = 5$ ), and CRF02\_AG-containing recombinants ( $n = 6$ ). The samples in the first subgroup reacted with seven to nine “homologous” and zero to two “heterologous” probes. In the second group, the lack of reactivity at the PCR level prevented the sample from hybridizing to the probes, as reflected in the lower number of



**FIG. 4.** MHAcrrf02 results on CRF02\_AG ( $n = 22$ ), CRF02\_AG-containing recombinants ( $n = 6$ ) and samples belonging to other clades ( $n = 34$ ). The number of samples reacting with different numbers of “homologous” and “heterologous” probes is shown. For the CRF02\_AG group, specimens with nonreactive PCRs are highlighted. The CRF02\_AG genomic content of second-generation recombinants is shown. See text for details.

reacting “homologous” probes observed (range = 7–5). In the CRF02\_AG-containing recombinants, a trend for samples with a lower content for CRF02\_AG in their genomic structure toward lower numbers of reacting “homologous” probes was observed.

Importantly, samples from all the other subtypes, CRFs, and unique recombinants clustered separately from the CRF02\_AG. In no case did any non-CRF02\_AG sample react simultaneously with more than three “homologous” and less than three “heterologous” probes. No clear identification of individual non-CRF02\_AG clades could be established based on the number of reacting “homologous” and “heterologous” probes, even though a trend could be observed (i.e., subtype A samples reacted with one to three “homologous” and four to five “heterologous” probes, subtype G reacted with four “homologous” and four “heterologous” probes, data not shown).

Thus, we could clearly establish a cut-off for the genotyping of a sample by MHAcrrf02: only samples that reacted simultaneously with four or more “homologous” and two or less “heterologous” probes would be classified as CRF02\_AG or CRF02\_AG-containing recombinants. Based on this classification, the MHAcrrf02 would perform with a sensitivity and specificity of 100%. The system proved to be so robust that neither the lack of reaction in some PCRs nor the recombination of CRF02\_AG with other clades interfered with the classification.

## DISCUSSION

The aim of present work was to design a genotyping tool to aid the evaluation of the HIV-1 epidemics in West/West Central Africa in the setting of large cohorts. The results presented in this paper clearly indicate that MHAcrrf02 meets the requirements for such a tool. First, MHAcrrf02 allows for the specific detection of CRF02\_AG and CRF02\_AG-containing recombinants in a milieu of other complex strains, including the parental subtypes A and G, other subtypes, CRFs, and URFs. None of the 34 non-CRF02\_AG samples assessed was classified as CRF02\_AG by this assay. High throughput, a crucial characteristic in tests aimed at the analysis of large cohorts, can also be achieved through MHAcrrf02 because of the use of the PCR and real-time PCR technologies.<sup>37</sup> With this assay, a trained technician having all the necessary resources could test ca. 100 samples in a week. In addition, MHAcrrf02 is applicable to primary clinical samples. The performance of the test on primary PBMCs was comparable to that observed on HIV-1 clones and cocultures (data not shown). When preceded by a retrotranscription step, the assessment of plasma HIV-1 RNA samples showed results equivalent to those of primary PBMCs (data not shown). Finally, MHAcrrf02 performed with high specificity and sensitivity.

MHAcrrf02, as any other genotyping method, has its limitations. The use of this assay is restricted to the study of the West/West Central African samples for which it was specifically designed. Other regional MHAs have been developed, such as MHAacd in East Africa,<sup>37</sup> MHAbf in South America,<sup>38</sup> or MHAbe in Southeast Asia.<sup>39</sup> Another limitation is HIV-1 within-clade genetic diversity, which may prevent, in some cases, the hybridization of primers and probes. In the current study, failure in the PCR amplification and probe binding was

observed, but in no case did it compromise the clear classification of the samples. Furthermore, using the criteria for identification proposed here, CRF02\_AG samples with up to 3/8 negative PCRs were also successfully identified. HIV-1 genetic diversity could also lead to the hybridization of probes to heterotypic strains, decreasing the specificity. In the present analysis, as the criteria of classification relied on the recognition of patterns of probe reactivity rather than the reactivity at a single or few loci, high specificity was achieved. The assessment of both “homologous” and “heterologous” probes allowed for the specific sorting out of CRF02\_AG strains be it from parental or from unrelated clades, and it rendered MHAcrrf02 a capacity to cope with the extensive HIV-1 genetic diversity found in the region. Moreover, the location of the probes at both sides of established recombination breakpoints in CRF02\_AG minimizes false positive results due to A/G intersubtype recombinants phylogenetically unrelated to the CRF.

The emergence of second-generation recombinants between CRF02\_AG and other clades is a phenomenon that should also be monitored.<sup>7,28,29</sup> These recombinants may share epitopes and other biologically relevant properties with the parental CRF, while losing some of the motifs that react with MHAcrrf02 “homologous” probes. The performance of this assay was tested on a set of second-generation recombinants with a CRF02\_AG genetic content ranging from 40 to 90%, and a varying level of loss of recognition by “homologous” probes was observed. However, 6/6 samples were successfully identified as CRF02\_AG-related, still reacting with a number of “homologous” and “heterologous” probes that clearly differentiated them from non-CRF02\_AG strains. Nevertheless, it is expected that strains with an even lower genetic content of CRF02\_AG may not be identified by this assay. The analysis of larger numbers of samples will provide a deeper insight into the performance of the assay.

The mosaic genomic structure of CRF02\_AG<sup>7,18</sup> on the one hand, and the level of complexity and extent of the HIV-1 epidemics in West/West Central Africa<sup>7,18</sup> on the other, seem to pose conflicting methodological requirements for genotyping tools. No other assay except for full-length genome (FLG) nucleotide sequencing, a low-throughput sophisticated technique, can be used to accurately establish the genotype of strains in this region of the world.<sup>7,18</sup> Cohort-based studies require the use of high-throughput assays to obtain adequate sampling of strains. The MHAcrrf02 assay can bridge this gap. MHAcrrf02 and FLG sequencing share the extensive examination of different areas of the HIV-1 genome, while the former has a much larger throughput. Nevertheless, MHAcrrf02 is not meant to replace FLG sequencing, rather it can be useful for the screening of large numbers of samples and for selecting either interesting specimens or a representative sample set for FLG sequencing.

Other approaches applied to this problem have included partial-genome sequencing<sup>5,6,8,10,12,33</sup> and HMA.<sup>25,34–36</sup> The latter, in its *gag/env* version, acknowledged the need to differentiate CRF02\_AG from subtype A and CRF01\_AE.<sup>35</sup> In addition to its higher throughput, MHAcrrf02 presents another advantage over partial sequencing and HMA: by assessing simultaneously eight different HIV-1 genomic regions, MHAcrrf02 is less prone to false positive results due to A/G recombinants unrelated to CRF02\_AG.

CRF02\_AG is recognized as one of the six clades most globally prevalent in the HIV-1/AIDS pandemic, together with sub-

types A, B, C, D, and CRF01\_AE.<sup>7,18</sup> It has been selected for use in candidate HIV-1 vaccines. This assay will permit more efficient characterization of the region, countries, and cohorts where CRF02\_AG is an important strain. Together with other regional genotyping assays, MHAcrf02 will contribute to the development of a global picture of the HIV-1 diversity and geographic distribution, providing a strong foundation for intervention. The first broad field test of this assay is underway.

## ACKNOWLEDGMENTS

The authors would like to thank the government of Cameroon for permission to undertake this study. We thank Dr. Michael Hoelscher for valuable discussion of the manuscript, Dr. Matthew Harris for sequences provision, N. Torimiro, U. Tamoufe, A.T. Prosser, A. Boupda, K. Long, L. Zekeng, S. Koulla-Shiro, B. Abissegue, L. Kaptue, M. Monny-Lobe, the Project Staff of the Hopkins Cameroon Project, and the U.S. Embassy in Yaounde for assistance in carrying out the work. The views and opinions expressed herein do not necessarily reflect those of the U.S. Army or Department of Defense. Support was received as part of a U.S. Department of Defense interagency agreement with Division of AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health, U.S. Department of Health and Human Services. This work was partially supported by an International Research Scientist Development award from the National Institutes of Health, Fogarty International Center (K01 TW00003-01 to N.W.).

## SEQUENCE DATA

Accession numbers for previously published Tanzanian, Kenyan, and Djibutian sequences used in the development of MHAcrf02 are AF361872, AF361873, AF457052, AF457063, AF457065, AF457066, AF457068, AF457069, AF457079, AF063223, and AF061640. GenBank accession numbers for novel sequences are AY371121–AY371170.

## REFERENCES

1. McCutchan FE: Understanding the genetic diversity of HIV-1. *AIDS* 2000;14(Suppl. 3):S31–44.
2. Robertson DL, Anderson JP, Bradac JA, *et al.*: HIV-1 nomenclature proposal. *Science* 2000;288:55–56.
3. Peeters M and Sharp PM: Genetic diversity of HIV-1: The moving target. *AIDS* 2000;14(Suppl. 3):S129–140.
4. Osmanov S, Pattou C, Walker N, Schwardlander B, and Esparza J: Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. *J Acquir Immune Defic Syndr* 2002;29:184–190.
5. Fonjungo PN, Mpoudi EN, Torimiro JN, *et al.*: Human immunodeficiency virus type 1 group m protease in Cameroon: Genetic diversity and protease inhibitor mutational features. *J Clin Microbiol* 2002;40:837–845.
6. Morison L, Buve A, Zekeng L, *et al.*: HIV-1 subtypes and the HIV epidemics in four cities in sub-Saharan Africa. *AIDS* 2001;15 (Suppl. 4):S109–116.
7. Carr JK, Torimiro JN, Wolfe ND, *et al.*: The AG recombinant IBNG and novel strains of group M HIV-1 are common in Cameroon. *Virology* 2001;286:168–181.
8. Ortiz M, Sanchez I, Gonzalez MP, *et al.*: Molecular epidemiology of HIV type 1 subtypes in equatorial Guinea. *AIDS Res Hum Retroviruses* 2001;17:851–855.
9. Nyambi P, Heyndrickx L, Vereecken K, *et al.*: Predominance of infection with HIV-1 circulating recombinant form CRF02\_AG in major Cameroonian cities and towns. *AIDS* 2002;16:295–296.
10. Fonjungo PN, Mpoudi EN, Torimiro JN, *et al.*: Presence of diverse human immunodeficiency virus type 1 viral variants in Cameroon. *AIDS Res Hum Retroviruses* 2000;16:1319–1324.
11. Zhong P, Burda S, Urbanski M, *et al.*: HIV type 1 group M clades infecting subjects from rural villages in equatorial rain forests of Cameroon. *J Acquir Immune Defic Syndr* 2002;31:495–505.
12. 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.
13. Ellenberger DL, Pieniazek D, Nkengasong J, *et al.*: Genetic analysis of human immunodeficiency virus in Abidjan, Ivory Coast reveals predominance of HIV type 1 subtype A and introduction of subtype G. *AIDS Res Hum Retroviruses* 1999;15:3–9.
14. Ellenberger DL, Li B, Lupo LD, *et al.*: Generation of a consensus sequence from prevalent and incident HIV-1 infections in West Africa to guide AIDS vaccine development. *Virology* 2002;302:155–163.
15. Andersson S, Norrgren H, Dias F, Biberfeld G, and Albert J: Molecular characterization of human immunodeficiency virus (HIV)-1 and -2 in individuals from Guinea-Bissau with single or dual infections: Predominance of a distinct HIV-1 subtype A/G recombinant in West Africa. *Virology* 1999;262:312–320.
16. 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.
17. Cornelissen M, van Den Burg R, Zorgdrager F, and Goudsmit J: Spread of distinct human immunodeficiency virus type 1 AG recombinant lineages in Africa. *J Gen Virol* 2000;81:515–523.
18. McCutchan FE, Carr JK, Bajani M, *et al.*: Subtype G and multiple forms of A/G intersubtype recombinant human immunodeficiency virus type 1 in Nigeria. *Virology* 1999;254:226–234.
19. Pandrea I, Robertson DL, Onanga R, *et al.*: Analysis of partial pol and env sequences indicates a high prevalence of HIV type 1 recombinant strains circulating in Gabon. *AIDS Res Hum Retroviruses* 2002;18:1103–1116.
20. Mamadou S, Montavon C, Ben A, *et al.*: Predominance of CRF02-AG and CRF06-cpx in Niger, West Africa. *AIDS Res Hum Retroviruses* 2002;18:723–726.
21. Sankale JL, Hamel D, Woolsey A, *et al.*: Molecular evolution of human immunodeficiency virus type 1 subtype A in Senegal: 1988–1997. *J Hum Virol* 2000;3:157–164.
22. Cham F, Heyndrickx L, Janssens W, *et al.*: Study of HIV type 1 gag/env variability in The Gambia, using a multiplex DNA polymerase chain reaction. *AIDS Res Hum Retroviruses* 2000;16:1915–1919.
23. Montavon C, Vergne L, Bourgeois A, *et al.*: Identification of a new circulating recombinant form of HIV type 1, CRF11-cpx, involving subtypes A, G, J, and CRF01-AE, in Central Africa. *AIDS Res Hum Retroviruses* 2002;18:231–236.
24. Takehisa J, Zekeng L, Ido E, *et al.*: Various types of HIV mixed infections in Cameroon. *Virology* 1998;245:1–10.
25. Toure-Kane C, Montavon C, Faye MA, *et al.*: Identification of all HIV type 1 group M subtypes in Senegal, a country with low and stable seroprevalence. *AIDS Res Hum Retroviruses* 2000;16:603–609.

26. Taniguchi Y, Takehisa J, Bikandou B, *et al.*: Genetic subtypes of HIV type 1 based on the *vpv/env* sequences in the Republic of Congo. *AIDS Res Hum Retroviruses* 2002;18:79–83.
27. 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.
28. Mamadou S, Vidal N, Montavon C, *et al.*: Emergence of complex and diverse CRF02-AG/CRF06-cpx recombinant HIV type 1 strains in Niger, West Africa. *AIDS Res Hum Retroviruses* 2003;19:77–82.
29. Janssens W, Salminen MO, Laukkanen T, *et al.*: Near full-length genome analysis of HIV type 1 CRF02.AG subtype C and CRF02.AG subtype G recombinants. *AIDS Res Hum Retroviruses* 2000;16:1183–1189.
30. 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.
31. Carr JK, Salminen MO, Albert J, *et al.*: Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G inter-subtype recombinants. *Virology* 1998;247:22–31.
32. 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.
33. 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.
34. Peeters M, Koumare B, Mulanga C, *et al.*: Genetic subtypes of HIV type 1 and HIV type 2 strains in commercial sex workers from Bamako, Mali. *AIDS Res Hum Retroviruses* 1998;14:51–58.
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. Study Group on Heterogeneity of HIV Epidemics in African Cities. *J Virol* 2000;74:363–370.
36. 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.
37. Hoelscher M, Dowling WE, Sanders-Buell E, *et al.*: Detection of HIV-1 subtypes, recombinants, and dual infections in east Africa by a multi-region hybridization assay. *AIDS* 2002;16:2055–2064.
38. Hierholzer J, Montano S, Hoelscher M, *et al.*: Molecular epidemiology of HIV type 1 in Ecuador, Peru, Bolivia, Uruguay, and Argentina. *AIDS Res Hum Retroviruses* 2002;18:1339–1350.
39. Watanaveeradej VCJ, Hoelscher M, Benenson M, Brown AE, Birs D, and McCutchan FE: High throughput screening assay for studying HIV genetic diversity in Asia. XIV International AIDS Conference. Barcelona, Spain, July 7–12, 2002. Abstract TuPeA4424.
40. Los Alamos National Laboratory HIV Sequence Database. Los Alamos, NM.
41. Felsenstein J: Phylip software package version 3.2c. Department of Genetics, University of Washington, Seattle, WA.
42. Maciukenas S: Treetool, 2.0.2. Ribosomal DNA Database Project, University of Illinois Board of Trustees, Urbana, IL, 1994.
43. Salminen MO, Carr JK, Burke DS, and McCutchan FE: Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses* 1995;11:1423–1425.
44. de Baar MP, Timmermans EC, Bakker M, de Rooij E, van Gemen B, and Goudsmit J: One-tube real-time isothermal amplification assay to identify and distinguish human immunodeficiency virus type 1 subtypes A, B, and C and circulating recombinant forms AE and AG. *J Clin Microbiol* 2001;39:1895–1902.
45. Kimura M: A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980;16:111–120.

Address reprint requests to:

Gustavo H. Kijak  
Henry M. Jackson Foundation  
1 Taft Court, Suite 250  
Rockville, Maryland 20850

E-mail: gkijak@hivresearch.org