Novel Multiregion Hybridization Assay for the Identification of the Most Prevalent Genetic Forms of the Human Immunodeficiency Virus Type 1 Circulating in Portugal

Ferdinando B. Freitas, Aida Esteves, João Piedade, and Ricardo Parreira

Abstract

The most efficient method for HIV-1 genetic characterization involves full-genome sequencing, but the associated costs, technical features, and low throughput preclude it from being routinely used for the analysis of large numbers of viral strains. Multiregion hybridization assays (MHA) represent an alternative for a consistent genetic analysis of large numbers of viral strains. Classically, MHA rely on the amplification by real-time PCR of several regions scattered along the HIV-1 genome, and on their characterization with clade-specific TaqMan probes (also known as hydrolysis probes). In this context, the aim of our study was the development of a technical variant of an MHA (vMHA_{B/G/02}) for genotyping the most prevalent genetic forms of HIV-1 circulating in Portugal. Different sets of primers were designed for universal and clade-specific amplifications of several sections of the viral genome: *gag*, *pol*(Pr), *pol*(RT), *vpu*, *env*(gp120), and *env*(gp41). vMHA_{B/G/02} was implemented using a real-time PCR-based approach, with detection dependent on the use of SYBR Green I. As an alternative, a technically less demanding strategy based on conventional PCR and agarose gel analysis of the reaction products was also developed. This method performed with overall good sensitivity and specificity (>91%) when a convenience sample of 45 plasma-derived HIV-1 strains was analyzed. Apart from the detection of subtype B, G, CRF02_AG, and CRF14_BG viruses, several unique B/G recombinant were also detected. Curiously, recombinant viruses including CRF02_AG sequences were not detected in the group of samples analyzed.

Introduction

THE EXTENSIVE GENETIC VARIABILITY that characterizes HIV-1 is reflected in the classification of viral strains into groups (M, N, O, and a putative group P),^{1,2} subtypes (A–D, F–H, J, and K), at least six subsubtypes (A1–A4 and F1–F2) in the pandemic group M, 50 circulating recombinant forms (CRF; HIV Sequence Database, <http://hiv-web.lanl.gov/> as of January/2012), and a profusion of unique recombinant viruses (URF; unique recombinant forms).

Over the years, the numerous HIV-1 genetic characterization studies published in the literature have been fueled by the potential impact of viral genetic variability on distinct biological properties of individual HIV-1 subtypes, which may translate into differences in transmission and disease progression rates and antiretroviral susceptibility, or a differential performance of diagnostic tests and viral load assays, as well as on vaccine design.^{1,3–5} Furthermore, the collected ge-

netic data have also seeded plentiful phylogenetic analyses of the circulating viruses, which have proven to be invaluable for epidemiological investigation, allowing the tracking of viral spread over space and time.^{6,7}

Until recently, the heteroduplex mobility assay (HMA)⁸ and sequence analyses of short segments of the viral genome (mostly from *gag*, *pol*, or *env*) were the most widely used experimental approaches for the genetic characterization of HIV-1. However, both methods have low genetic resolution since only a partial inspection of the viral genomic makeup is carried out. Alternatively, high genetic resolution may be achieved with full-length genome sequence analysis. Nevertheless, its low-throughput, necessary high analytical expertise, and relative high cost still limit its usefulness, especially when large cohorts are being assessed. Fortunately, these drawbacks have been partially overcome with the development of genotyping assays collectively known as multiregion hybridization assays (MHA).

In their original setup MHA are based on the amplification, in a first step, of multiple viral genomic regions (usually ≥ 5), regardless of viral genotype, using so-called universal primers. The obtained amplicons are then split into a series of second rounds of real-time polymerase chain reactions (PCR) where they serve as template, and during which the fluorescence emitted by clade-specific fluorescent probes allow the selective identification of specific viral genotypes. MHA are particularly useful to describe complex epidemics since they can be designed to account for the viral genetic diversity in specific areas of the globe.^{9–17} Furthermore, this kind of approach also provides valuable information on the relative proportion of circulating strains, identifies new molecular viral forms as they arise, and allows the recognition of putative dual infections when different clade-specific probes hybridize to a single genomic region.

The objective of this work was to design a variant of an MHA-type assay (vMHA) for the characterization of the HIV-1 molecular forms circulating in Portugal, where the epidemic is, up to the present, dominated by subtype B and G, as well as CRF02_AG, CRF14_BG^{18–21} and unique recombinant strains that include subtype B, G, and CRF02_AG sequences. This assay (vMHA_{B/G/02}) has been tailored to work both in a real-time PCR format using *SYBR Green I* as the detection agent, or as a classical PCR/gel electrophoresis based approach, which allows its implementation in a wide range of laboratory settings.

Materials and Methods

Samples

The initial evaluation of the performance of the amplification primers designed in the course of this work was carried out using seven HIV-1 reference strains. A whole cell extract of the 8E5/LAV cell line (a derivative of A3.01 cells containing a single integrated copy of proviral DNA coding for defective viral particles) and pNL4-3 (full-length replication and infection-competent chimeric DNA clone) were used as references for subtype B viruses (obtained from the Centralized Facility for AIDS Reagents, National Institute for Biological Standards and Control, UK). A set of four pGEM-T Easy (Promega, USA) derivatives carrying full-length proviral HIV-1 genomes (PT2695–AY612637, PT3037–FR846408, PT3306–FR846409, PT988–FR846410) (unpublished; direct submission to the GenBank/EMBL/DDBJ databases) and pBD6-15²² were used, respectively, as subtype G and CRF02_AG references. Except for 8E5/LAV, all the other HIV-1 references were used as purified plasmid DNA, extracted from *Escherichia coli* hosts using the QIAGEN MIDI kit (QIAGEN, Germany).

The performance of the vMHA_{B/G/02} assay was assessed on a panel of 45 clinical samples (plasma), collected from HIV-1-seropositive individuals living in the Lisbon (Portugal) metropolitan area. Different regions of the proviral genome of HIV-1 strains present in these samples were previously subtyped by HMA and/or sequencing analysis of the genes *gag* and/or *env*^{18,19} and *nef*.²¹

cDNA synthesis and PCR amplification

Viral RNA was extracted from human plasma using the Instant Virus RNA kit (Analytik Jena, Germany). Briefly, 150 μ l of plasma was added to 450 μ l of lysis solution (RL), and

this mixture was homogenized by vortexing. After 15 min at room temperature, 600 μ l of binding solution (RBS) was added. This new mixture was again homogenized by vortexing, applied onto a spin filter, and quickly centrifuged at approximately $10,000\times g$, for 1 min. After two washes with the washing solutions provided (*HS* and *LS*), the extracted RNA was eluted in 60 μ l of RNase-free water, added 30 μ l at a time, followed by incubation at room temperature (2 min). The eluate containing the extracted RNA was split into three different aliquots with volumes of 5 μ l, 5 μ l, and 20 μ l, respectively. One of the 5- μ l aliquot was used for reverse transcription (RT) while the others were stored at -80°C .

Reverse transcription of viral RNA was carried out with the RevertAid H Minus First Strand cDNA Synthesis kit (Fermentas, Lithuania). In sum, 5 μ l of extracted RNA was added to a reaction mixture consisting of 1 μ l of random hexamer primers and 6 μ l of diethylpyrocarbonate (DEPC)-treated water. This mixture was incubated at 70°C for 5 min (for RNA denaturation) and then placed on ice for an additional 5 min. A second reaction mix containing 4 μ l of reaction buffer, 1 μ l of Ribolock RNase Inhibitor, 2 μ l of dNTP Mix (10 mM), and 1 μ l of reverse transcriptase enzyme (RevertAid H Minus M-MuLV Reverse Transcriptase, in a concentration of 200 U/ μ l) was prepared, and later added to the first. Negative controls were included in each set of RT reactions in order to assess for DNA contaminations.

The obtained cDNA served as a template for the amplification of HIV-1 sequences throughout this work using the PuRe Taq Ready-to-Go PCR Beads (GE Healthcare, UK). Overall, the generated amplicons varied in size, between (approximately) 200 and 1,000 nucleotides (nt), depending on the region analyzed. However, for some samples, and in particular for the gp120 coding region, the obtained amplification products were no larger than 120 nt. The PCR primers used in the course of this work are listed in Table 1. Most the oligonucleotides used present a G+C content ranging from 30% to 60%, a size varying from 20 to 30 nt, and a melting temperature usually not exceeding 65°C .

vMHA_{B/G/02} and genome targets

The principle of the MHA assays in their classical format rests on the amplification of short fragments along the HIV-1 genome, and the detection of clade-specific fluorescent probe hybridization in a real-time PCR setup. vMHA_{B/G/02} was developed to detect HIV-1 subtypes B and G and their recombinants (e.g., CRF14_BG) as well as CRF02_AG sequences, and was implemented in a nested/heminested format, differing from the classical MHA formulation as a result of (1) the introduction of an additional universal amplification step (useful for low viral load samples, ≤ 50 copies/ml of plasma, which, otherwise, normally remain undetected) and (2) the use of clade-specific PCR primers (and not real-time PCR hydrolysis probes) for viral genotyping. In separate first-round PCR reactions, five regions of the HIV-1 genome [*gag*, *pol*(Pr), *pol*(RT), *vpu*, *env*(gp120), and *env*(gp41), see Fig. 1] were amplified with universal primers (boldface in Table 1). Each amplicon was then usually used as a template for second-round PCR amplifications with universal primers, followed by a third amplification step with one clade-specific (one for each subtype, B, G, or CRF02_AG) and a universal oligonucleotide (schematically represented in Fig. 2 as the

TABLE 1. UNIVERSAL (BOLDFACE) AND CLADE-SPECIFIC PRIMERS USED IN THIS WORK

Target region(s)	Primer designation	Sequence (5' → 3')	Coordinates ^a	Orientation
<i>gag</i>	GagOF	TGGATGACAGAMACCTTGYTGGTCC	1735–1759	Forward
	GagOR	CTTCTAAYACTGTATCATCTGCTCCTGT	2328–2355	Reverse
	GagIF	TAGAAGAAATGATGACAGCATGYCAG	1817–1842	Forward
	GagIR	CCYTCCTTYCCACATTTCCAACAGCC	2023–2048	Reverse
	GagB	GGCTGAAGCAATGAGCCAAGTAACAAAT	1878–1905	Forward
	Gag02	GGCYGAGGCAATGAGTCAAGYACAACAR	1878–1905	Forward
	GagG	AGCTGAGGCAATGAGCCWGGCATCAGGK	1880–1907	Forward
<i>pol(Pr)</i>	PrRtOF	ACAGGAGCAGATGATACAGTRTTAGAAG	2328–2355	Forward
	PrRtOR	TTCCATCCYTGTGGAAGCACATTGTACTGATA	2979–3010	Reverse
	PrRtIF	ATGGAARCCAAAAATGATAGGGGGAATTGG	2375–2404	Forward
	PrRtIR	TGGRCCATCCATTCTGGCTTTAAATT	2581–2606	Reverse
	PrRtB	CARATACYCATAGAAATCTGYGGACATAAA	2433–2462	Forward
	PrRtG	CARATACTTATAGAAATTTGTGGA AAAARG	2433–2462	Forward
<i>pol(RT)</i>	RTOF	CAGTACAATGTGCTTCCACARGGATGG	2982–3008	Forward
	RTOR	TAAYTGYTTTACATCATTAGTGTGRGC	3627–3653	Reverse
	RTIF	GCTGGACTGTCAAYGAYATACARAART	3301–3327	Forward
	RTIR	TCTTGATARATTTGRTATGTCCAYTGG	3554–3580	Reverse
	RTB	AGTAATACCACTAACAGAAGAAGCAGAGCTA	3422–2452	Forward
	RT02	TATAGTAMCACTGACTGAGGAAGCAGAATTA	3422–2452	Forward
	RTG	CATAGTAYCACTRACWGCAGAAGCAGAATTG	3422–2452	Forward
<i>vpu-env(gp120)</i>	VprOF	GATGGAACAAGCCCCAGARGACCARGG	5558–5584	Forward
	Gp120OR	CACATGGYTTTAGGCTTTSRTCCCATATA	6556–6582	Reverse
	VprIF	CTTAGGCATCTCCTATGGCAGGAAGAAGC	5956–5984	Forward
	Gp120IR	CATTKCCACTRTCTTCTGCTCTTTC	6203–6227	Reverse
	VpuB	CTATCAAAGCAGTAAGTAGTAYATGT	6035–6060	Forward
	VpuG	GTACCAAAGCAGTRAGTARTATAAATTAR	6035–6053	Forward
	Gp120IF-1	GGGTCACRGTCTAYTATGGRTACCTGTG	6328–6356	Forward
	Gp120IR	GGGYTRGGGTCTGTGGGTACACAGGCA	6440–6466	Reverse
	Gp120B-1	TGCACAAAATAGAGTGGTRGTGCTTCYTCC	6358–6389	Reverse
	Gp120B-2	AGAGTGGTGGTTGCTTCATT	6360–6379	Reverse
	Gp12002-1	TAGACCGTGACCCACAAMTYTTCAGCMT	6310–6340	Reverse
	Gp12002-2	CGTGACCCACAAMTYTTCAG	6315–6335	Reverse
	Gp120G-1	GTGTAGCCCAGACATTATGGYTTTCAGTACT	6408–6438	Reverse
	Gp120G-2	CCCATACATTATGGYTGTCTAGTACT	6414–6438	Reverse
<i>env(gp41)</i>	Gp41OF	AATAGAGTTAGGMAGGGATAYTCACC	8340–8365	Forward
	NefOR	TGGCCCTGGTGTGTARTTCTGCCAATC	9163–9189	Reverse
	Gp41IF	CTGTGYCTCTTCAGCTACCACCGMTTG	8511–8537	Forward
	Gp41IR	CTRTCTGTCCMVTYAGCTACTRCTAT	8682–8707	Reverse
	Gp41B	CTTCTGGGACGCAGGGGGTGG	8574–8594	Forward
	Gp41G	CAGCAGYCTCAAGGGACTGAGACT	8587–8589	Forward

^aPrimer coordinates are relative to the HIV-1 subtype B reference sequence HXB2 (accession number K03455).

three-step protocol). Alternatively, each amplicon from the first-round amplification reaction was used directly for the clade-specific amplification (Fig. 2, two-step protocol).

For the reference samples, each region of the viral genome was amplified in a total volume reaction of 25 μ l using the PuRe Taq Ready-to-Go PCR Beads (GE Healthcare, UK), 5 μ l input DNA, and 400 nM of each universal primer.

Universal PCR amplifications were performed using cycling conditions that included an initial denaturation step at 94°C for 1 min, followed by 35 cycles of 94°C for 45 s, 55°C for 45 s [although they were tested over a wide range of hybridization temperatures (HT): 55°C, 58°C, 60°C, 63°C, 66°C, and 69°C], and 72°C for 1 min. Conditions for the clade-specific amplification of viral sequences were similar, except that (1) the hybridization temperature that varied according to the targeted region and (2) the DNA input that was lower than 1 μ l.

Real-time PCR amplifications were carried out in a Rotor-Gene 3000 thermocycler (Corbett Research, UK) using reaction volumes of 12.5 μ l, 800 nM of each primer, and the Maxima SYBR Green qPCR Master Mix 2x (Fermentas, Lithuania) allowing detection of an amplification product depending on the emission of fluorescence by DNA-bound SYBR Green I. Positive amplification results were all those for which fluorescence intensity increased exponentially over, at least, five consecutive cycles, and a cycle threshold (C_t) < 30. An analysis of the melting curves of the obtained PCR products was systematically carried out.

For each region of the viral genome for which different types of specific amplification primers were defined, let “a” be the number of amplifications where probe and target were homologous, “b” the number of amplifications where a homologous probe failed to perform (i.e., a specific primer did not allow the amplification of a homologous target), “c” the number of off-target amplifications (i.e., the subtype of the

FIG. 1. Schematic representation of the different HIV-1 genomic targets analyzed with vMHA_{B/G/02}. The white rectangles indicate the different coding and regulatory regions of the HIV-1 proviral genome (downloaded from www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html), identified by standard abbreviations. The small number in the upper left corner of each rectangle indicates the position of the A in the ATG start codon, while the number in the lower right records the last nucleotide of the translation stop codon. These coordinates are those of the HXB2 subtype B reference sequence (accession number K03455). Light gray boxes indicate first-round universal amplicons, dark gray/black hybrid rectangles indicate second-round universal amplicons, where the section amplified by the clade-specific primers is restricted to the black portion of the rectangle. The orientation of the clade-specific primers (designations indicated in the figure as well as in Table 1) is that indicated by the arrows.

probe and target sequence did not match), and “d” the number of failed amplifications when the subtype of the probe and sequence did not match. Under these circumstances sensitivity was defined as $a/(a+b)$ whereas specificity reflected the ratio $d/(c+d)$.

DNA sequencing and phylogenetic analysis

PCR products were purified with QIAquick columns (QIAGEN, Germany) prior to DNA sequencing. Phylogenetic

analyses of viral sequences were based on the construction of multiple sequence alignments using MAFFT-6,²³ from which corrected genetic distance matrices were obtained using Kimura’s two-parameter formula and MEGA-4.²⁴ Neighbor-joining phylogenetic trees included reference sequences taken from the Los Alamos HIV sequence database (www.hiv.lanl.gov/content/sequence/HIV/mainpage.html). Bootstrap analyses were performed with 1,000 resamplings of the original sequence data. The putative mosaic structure of the sequences analyzed was investigated by bootscanning using SimPlot 3.5.1.²⁵

FIG. 2. vMHA_{B/G/02} two-step vs. three-step protocol. In a first set of PCR reactions, universal primers were used to amplify different regions of the viral genome, regardless of their genetic identity. In the two-step protocol, the first-round amplicons were split into independent amplification reactions set with both clade-specific and universal primers [forward (Fwd) and reverse (Rev) orientations, respectively, in the figure]. In the three-step protocol, following a first universal amplification with external (Ext) primers, a second nested universal PCR reaction will follow using internal (Int) oligonucleotides. The amplification makes use of a clade-specific/universal primer combination, as previously stated.

FIG. 3. Competition assays. Representative example of an evaluation of the specificity of clade-specific *env(gp41)* primers used for subtype B amplification by real-time PCR in the presence of a subtype G (A) or CRF02_AG (B) competitor template in a 1:1, 10:1, and 100:1 ratio (competitor/specific DNA).

The DNA sequences obtained in the course of this work were deposited in the GenBank/EMBL/DDBJ databases under the following accession numbers: *gag* (FR848963–FR849006), *pol(Pr)* (FR850059–850102), *pol(RT)* (FR850289–FR850330), *vpu* (HE583228–HE583271), *env(gp120)* (FR870327–870369), and *env(gp41)* (FR870477–870518).

Results

Primer/probe design

The MHA assay here described was developed to allow a relatively rapid and unambiguous identification of the most prevalent HIV-1 variants that characterize the Portuguese HIV-1 epidemic, i.e., subtypes B and G (and their recombinants) as well as CRF02_AG.^{18–21} As in any of the MHA formulations previously described in the literature,^{9–17} this assay was conceived so as to allow the assignment of a genetic identity to several ($n=6$) short fragments scattered throughout the HIV-1 genome, revealing insights into the overall genetic makeup of viral strains without having to make use of whole-genome sequencing. The selected targets of the vMHA_{B/G/02} assay included sections of the *gag* (targeting the *p24/p7* region), *pol* (*Pr* and RT coding sequences), *vpu*, and *env* (*gp120* and *gp41*) coding sequences (Fig. 1).

The different amplification primers, whether universal or clade-specific, were designed using, as a guide, a multiple sequence alignment of 57 HIV-1 complete genomes downloaded from the Los Alamos HIV sequence database. This panel of reference sequences included subtypes A ($n=2$), B

($n=19$), C ($n=1$), F ($n=2$), G ($n=12$), H ($n=1$), K ($n=1$), J ($n=1$), plus recombinant forms CRF02_AG ($n=11$) and CRF14_BG ($n=7$), and was chosen to virtually cover the genetic diversity of HIV-1 already found in Portugal.^{18–21}

The complementary sequences to the universal primers (both inner and outer) were selected as regions of high inter-subtype conservation, making sure they flanked the known recombination breakpoints previously mapped on the CRF02_AG and CRF14_BG genomic structure (www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html). The clade-specific amplification primers used were those complementary to sections of high intraclade conservation but with high intersubtype divergence. Taking into account the different regions of the viral genome chosen as targets, as well as the sequence features of the subtypes and CRF of HIV-1 that we sought to identify (B, G, CRF02_AG, and CRF14_BG), different numbers of clade-specific primers were designed for the different viral genomic regions. Therefore, for the *gag*, *pol(RT)*, and *env(gp120)* regions, three specific primers were planned, allowing distinction between viruses of subtypes B and G/CRF14_BG and CRF02_AG sequences. For the *pol(Pr)*, *vpu*, and *env(gp41)* regions two specific primers were defined, allowing distinction between viral sequences of subtypes B and G/CRF02_AG/CRF14_BG. For the *env(gp120)* region, two sets of alternative clade-specific primers were designed (suffix -1, -2, Table 1).

The inclusion of degenerate nucleotide positions in the primers sequences was kept to a minimum, and was limited to the cases when it was necessary to increase their universality

FIG. 4. vMHA_{B/G/02} results for 45 clinical samples from HIV-1-seropositive intravenous drug users (IDUs) living in the Lisbon metropolitan area. The genomic identity of each target was determined by phylogenetic analysis of DNA sequences (as indicated in the Materials and Methods section). The overall genetic structure of the subtype B, subtype G, CRF02_AG, and CRF14_BG HIV-1 strains is indicated above each group of samples analyzed. The thin black rectangles indicate the approximate location on the viral genome of each of the targets analyzed. The two sets of genotyping results using the alternative *env(gp120)* primer sets (-1 and -2) are indicated within the thin-lined vertical rectangles. Whenever the quality and/or complexity of a given sequence precluded its unambiguous genotyping by phylogenetic analysis, the results obtained with vMHA_{B/G/02} are indicated. Expected/unexpected amplification results were defined as vMHA_{B/G/02} genotyping results that agreed/did not agree with those obtained using phylogenetic analysis, respectively. Multiple amplifications refer to multiple amplifications of a given viral target that result from hybridization of multiple sets of clade-specific primers to it.

and expected specificity. In a few cases, mismatches were introduced in the primer sequence so as to disfavor heterologous hybridizations relative to the homologous ones (see the following section). The primers used in this work were designed to present, as much as possible, a melting temperature (T_m) below 65°C, and include up to three nucleotide C/G at their 3'-end for increased specificity.^{26,27} However, for some regions of the viral genome analyzed, the melting temperatures were set to be higher than 65°C as a result of either a high C/G content or overall intersubtype sequence conservation of the target region [e.g., *env(gp41)* and *gag*, respectively].

Assay performance using a set of reference samples

The initial definition of PCR conditions that maximize the performance of all the universal primers used was set up against a set of reference HIV-1 genomes. The temperature range used in the hybridization step of the PCR protocols (55°C, 58°C, 60°C, 63°C, 66°C, and 69°C) did not seem to affect the specificity of the universal amplification reactions. In fact, a specific amplicon was obtained for each homologous primers/template DNA combination used, for most of the HT tested. However, an increase in HT did impact the amount of obtained PCR product, and for this reason, all the universal amplification reactions were further carried out at 55°C. A similar range of HT was tested for the clade-specific amplification reaction of the different target regions. Specific amplicons could be easily identified either by the analysis of aliquots of the reaction mixtures in agarose gels, or using a real-time approach for the intended amplifications with the following HT profile: *gag*–69°C, *pol(Pr)*–58°C, *pol(RT)*–60°C, *vpu-env(gp120)*–58°C, and *env(gp41)*–63°C.

Two alternative sets of clade-specific primers (suffixes -1 and -2 in Table 1) were defined in the reverse orientation for the analysis of the *env(gp120)* region. Primers -1 and -2 had a very similar sequence, with primers -1 being slightly larger than primers -2. Deliberate mismatches were introduced in the sequence of the -2 primer set, which were meant to deter heterologous hybridizations more strongly than they would affect the homologous ones.

Competition assays

We sought to assess the binding specificity/sensitivity of the clade-specific amplification primers to their homologous target, and how this could be influenced by the presence of a competitor (heterologous) DNA at concentrations similar to, or higher than (1:1, 10:1, 100:1) that of the homologous counterpart in the reaction mixture. This was carried out by performing competition assays. Since the specific amplifications described in the previous section showed similar results for all samples of the same subtype/CRF, the competition assays were carried out using one chosen representative of each subtype/CRF test (pNL4-3 for subtype B, p2695 for subtype G, pBD6-15 for CRF02_AG). For convenience, these assays were carried out using a real-time PCR format where the presence of an amplification product was revealed by the increase in fluorescence due to the presence of *SYBR Green I* in the reaction mixtures. The hybridization temperatures used were the ones defined in the previous section. The analysis of amplification curves showed that for all the six targeted re-

gions of the HIV-1 genome, the presence of a heterologous competitor DNA in concentrations up to 100 times higher than those of the homologous sequence did not influence the sensitivity/specificity of binding of the clade-specific primers to their homologous targets (see Fig. 3 for an example). In fact, similar amplification results were obtained in the absence of competitor DNA, using appropriate dilutions of the homologous DNA as template (data not shown). Finally, the analysis of the obtained melting temperature curves revealed the presence of a single amplicon in all PCR reactions (data not shown).

vMHA_{B/G/02} validation and genotyping results

Although the use of a panel of reference HIV-1 genomes allowed the definition of general PCR amplification conditions for the different sets of universal or clade-specific viral targets under study, the performance of vMHA_{B/G/02} was assessed against a panel of plasma samples (as previously defined) obtained from HIV-1-seropositive intravenous drug users (IDUs).

Unlike the HIV-1 references used, the template for the first amplification reactions corresponded to cDNA retrotranscribed from viral RNA extracted from human plasma, as other authors have previously shown that RNA-MHA is more sensitive than DNA-MHA.⁹

All the universal amplifications were carried out using 55°C as the HT. The specific DNA fragments with expected size were firstly amplified in a nested PCR approach with two sets of universal primers (three-step protocol, Fig. 2). This strategy was proven valuable, especially when working with low viral load. However, for a randomly selected group of samples, encompassing 20% of those under analysis, the expected clade-specific DNA fragments were amplified with a single set of universal primers (two-step protocol, Fig. 2).

Under the amplification conditions that were defined using the panel of HIV-1 references, the majority (39/45) of the *gag(p24/p7)* originated unambiguous results, while for the other genomic regions a significant number of nonspecific amplifications was obtained (multiple amplification profiles, data not shown). These were not totally unanticipated due to the small number of HIV-1 references used, the genetic diversity of which was expected to be considerably smaller than that of the clinical samples used. For this reason, a fine-tuning of the specific amplification conditions was carried out with a randomly selected subset of the clinical samples. This optimization of the PCR protocol involved successive increases (increments of 1°C) in the HT used for each clade-specific amplification, up to values for which unambiguous genotyping profiles could be defined: *gag*–69°C, *pol(Pr)*–63°C, *pol(RT)*–61°C, *vpu*–63°C, *env(gp120)*–66°C, and *env(gp41)*–68°C.

The combined results of vMHA_{B/G/02} genotyping and phylogenetic analyses are schematically represented in Fig. 4. An example of the phylogenetic analysis for the *pol(Pr)* region is shown in Fig. 5 (a similar approach was used for all the other study regions). Thirteen samples seemed to carry HIV-1 subtype B viruses (28.9%) while 14 other (31.1%) were shown to have subtype G viruses. Viral strains with a genetic structure compatible with CRF14_BG and CRF02_AG were identified, respectively, in six (13.3%) and four (8.9%) samples. Finally, a group of eight samples (17.8%) seemed to carry a multitude of unique B/G recombinant viruses.

FIG. 5. Phylogenetic analysis for the *pol(Pr)* genomic region. Some of the sequences obtained for this region during this work are highlighted with a square clustering with the respective subtype of the reference sequences.

Calculations of sensitivity and specificity of the method relied on the assignment of a genotype/genetic structure (in the case of mosaic genomes) to the viral targets, which was accomplished, respectively, by phylogenetic/bootscanning nucleotide sequence analyses. The sensitivity of the primers used were 98.0%–*gag*(p24/p7), 91.9%–*pol*(Pr), 94.4%–*pol*(RT), 97.1%–*vpu*, 97.3%–*env*(gp120, set 1), 94.4%–*env*(gp120, set 2), and 95.2%–*env*(gp41). The main factor limiting the sensitivity of the clade-specific primers used was their failure to hybridize, and consequently to amplify their targets. Detection of a specific PCR amplification by real-time PCR (as revealed by the analysis of melting curves) was more sensitive than the analysis of PCR reaction products by agarose gel electrophoresis, and if only this method was taken into account the corrected sensitivities were as follows: 98.0%–*gag*, 91.9%–*pol*(Pr), 98.1%–*pol*(RT), 97.1%–*vpu*, 98.4%–*env*(gp120, set 1), 94.4%–*env*(gp120, set 2), and 96.2%–*env*(gp41).

The same primers performed with the following average specificities when both methods of PCR amplification/detection were considered: 90.6%–*gag*, 85.2%–*pol*(Pr), 95.1%–*pol*(RT), 91.3%–*vpu*, 83.8%–*env*(gp120, set 1), 85.0%–*env*(gp120, set 2), and 98.1%–*env*(gp41). When only the real-time PCR results were used, the specificity values obtained were as follows: 93.8%–*gag*(p24/p7), 92.3%–*pol*(Pr), 97.7%–*pol*(RT), 94.1%–*vpu*, 83.8%–*env*(gp120, set 1), 85.0%–*env*(gp120, set 2), and 98.1%–*env*(gp41). The major limitation of specificity was the hybridization of primers to both homologous and heterologous target sequences (multiple amplification results), this being particularly evident in the *env*(gp120) region of viruses whose genotyping results suggested a structure compatible with the mosaic defined for CRF02_AG genomes.

Discussion

The extreme plasticity of the HIV-1 genome, and the potential for the generation of new recombinant viruses that may spread when opportunity arises, continuously present potential new challenges for diagnosis, treatment, and prevention.⁶ Over the years, most genetic characterization studies of circulating HIV-1 strains in a given population and/or geographic region have been carried out with methods that disclosed a very partial image of the genetic structure of the viral genomes. This was due to the fact that despite their relatively high throughput, the experimental approaches used (usually based on heteroduplex mobility assay or partial sequence analysis) have been focused on the study of a single, or a small number of viral genomic segments. Surely, whole genome sequencing stands as the most robust approach for a thorough characterization of the viral genome. However, its low throughput, high cost, demanding technical/analytical features, and associated computational resources preclude it from being the method of choice for a systematic analysis of HIV-1 genomes in most laboratory settings.

MHAs have been developed to bridge the gap between the characterization of each HIV-1 strain by sequencing of the entire genome and the requirement for adequate sampling of large and complex epidemics. Furthermore, they represent a possible solution to the growing problem of HIV-1 genotyping of epidemics characterized by the cocirculation of multiple viral subtypes and recombinant forms. Up to the present, a series of MHA designed to allow the characterization of regional epidemics include the MHAcd for East Africa,^{10,13}

MHAcrf02 for West/West Central Africa,¹⁴ MHA_B for South America,¹² and two versions of MHA_Bce (vs1 and vs2) for Asia.^{15,17} Accordingly, the MHA described in this work (vMHA_{B/G/02}) was planned to allow a rapid and unambiguous discrimination of the most prevalent HIV-1 subtypes (subtype B and G) and circulating recombinant forms (CRF02_AG and CRF14_BG) that comprise the majority of the HIV-1 strains circulating in Portugal.^{18–21} In a recent study²⁸ the same predominant subtypes were found, although an unusually high (and possibly nonrepresentative) prevalence of subtype C was reported. However, the basic features of your method allows for quick updating in the face of a change in distribution of subtypes in Portuguese epidemiology, such as an increased prevalence of a nondominant subtype.

All the MHAs that have been described in the literature follow a general two-step protocol. In a first round of PCR, different viral targets are amplified, irrespective of their genetic nature, with so-called outer universal primers. In turn, the obtained amplicons serve as a template for the subsequent amplification of viral sequences using inner universal primers in a real-time PCR setting, in which the specificity of the genotyping reaction derives from the inclusion of a clade-specific TaqMan probe in each of the different reaction mixtures. The hybridization of these probes to their homologous targets, and their subsequent degradation by the moving polymerase with emission of fluorescence, ensure the specificity of the subtyping result.

The MHA methodology here presented is a variation of the original protocol, and was developed to allow HIV-1 subtyping even in a laboratory setting with (1) limited/no access to a real-time PCR thermocycler and/or (2) limited financial resources for the acquisition of the expensive fluorescent hydrolysis probes. The combination of universal and clade-specific primers under PCR conditions optimized for different regions ($n=6$) of the HIV-1 genome resulted in the identification of clade-specific amplicons by conventional agarose gel electrophoresis. In this study, a three-step protocol was also tested, with two consecutive universal amplifications followed by a clade-specific one. When compared to the classical MHA protocol, the introduction of an additional amplification step might be viewed as an unwelcome increase in the experimental workload. However, it does raise the possibility of obtaining genotyping results even when working with samples with a very low viral load (≤ 50 copies/ml of plasma) from which viral-specific amplicons are difficult to obtain.²⁹ Furthermore, because the consecutive universal amplifications follow a nested PCR protocol, this results in an increase in both the sensitivity and specificity of the method. In case a real-time thermocycler is available, the method was shown to work with samples with low viral load with a detection step involving the use of SYBR Green I.

The main factor that seemed to limit the sensitivity of the clade-specific PCR amplifications was the failure of the clade-specific primer to hybridize to its homologous target. Although all the primer pairs used showed overall good sensitivity (91–98% range), our results also indicated that a single mismatch introduced at the 3'-end of a subtype-specific PCR primer was sometimes sufficient to deter a homologous hybridization under a given set of PCR conditions (data not shown). Similarly, relatively high specificity results characterized the different clade-specific amplifications (83–98% range), especially in the context of vMHA_{B/G/02} genotyping

using a real-time PCR setting. Multiple target hybridization (hence multiple PCR amplification) was, without a doubt, the major limitation of the assays' specificity. This was particularly evident in the case of the *env(gp120)* genomic region of viruses with putative CRF02_AG genomes, for which an almost systematic multiple hybridization profile was obtained, regardless of which alternative primer set (set 1 or set 2) was used.

Although a considerable number of CRF02_AG sequences (19.3%, 11/57) were included in the multiple sequence alignment that initially assisted primer design, the analysis of the obtained DNA sequences showed an unexpected sequence diversity in the *env(gp120)* amplicons of putative CRF02_AG sequences. This fact may contribute to lowering the specificity of hybridizations of multiple sets of primers to this target. Nevertheless, a slightly lower sensibility and specificity of the method here described, when compared to the traditional MHA assays, were expected. Traditionally, the performance of MHA assays is evaluated against a panel of reference samples, previously characterized from a genetic point of view. Since these samples serve both as a starting point for primers/probe design and as the reference material against which the primers' performance is tested, it is not unusual that very high sensitivity/specificity values are attained. In the case of the method here described, all the primers were designed taking into account a virtual alignment of sequences withdrawn from the public databases, carefully chosen to represent the diversity of subtypes/genetic forms of HIV-1 circulating in Portugal. The performance of the method was directly evaluated using clinical samples where a high HIV-1 diversity is expected, and which certainly contributed to an apparent decrease in the methods' sensibility and specificity vis-à-vis the traditional approach.

With some primer combinations, multiple amplification results were obtained, which also affected the performance of the alternative sets of clade-specific PCR primers designed for the analysis of the *env(gp120)* region. Although multiple sequence hybridization may be directly affected by high sequence variability, it may also disclose a mixed viral infection.

Frequently, mixed infections are revealed using a strategy that combines cloning of PCR products and the arbitrary selection of plasmid clones, followed by extensive DNA sequence analysis. Although it has previously been used for the clarification of dual/discordant hybridization results in the context of MHA analysis,¹³ this method is potentially unfruitful if (1) clone sampling is not extensive enough or (2) one virus is present at very low level. Alternatively, mixed infections can be tracked using a simpler approach, based on the identification of heteroduplexes formed by autologous PCR product hybridization, and their analysis.^{30,31} Although the method ultimately relies on DNA sequencing, it does allow the inspection of larger numbers of samples by combining HMA and colony PCR.

Despite being similar in their sequence, set 1 and set 2 (Table 1) of *env(gp120)* primers did not perform exactly alike. Although either set, at times, gave rise to multiple amplification products that could not be differentiated in a real-time PCR setup (using *SYBR Green I*), they were, indeed, distinguishable by agarose gel analysis. Although this suggests a strong impact of the high *env(gp120)* variability on the vMHAB/G/02 specificity, clearly only in-depth genotyping would help to distinguish between the impact of sequence

variability or a mixed infection in the performance of the method here described. A technological upgrade of vMHAB/G/02 might ensue from the use of clade-specific hydrolysis probes in PCR mixtures, as all other MHA suggest.

In this case, however, the assay (or any variant developed to deal with other HIV-1 epidemics) would be restricted to a laboratory setting with a real-time thermocycler and financial resources to acquire the relatively expensive fluorescent-labeled hydrolysis probes. Furthermore, this upgrade would require not only a reevaluation of the clade-specific oligonucleotide sequences to perform as fluorescent probes but also PCR conditions used for the amplification reactions. Indeed, TaqMan probes matching the exact sequence of the *env(gp41)* clade-specific primers listed in Table 1 showed maximum specificity (100%, for both B and G sequences) but considerably lower sensitivity (data not shown).

The subtyping results obtained with vMHAB/G/02 (subtype B–28.9%, subtype G–31.1%, CRF14_BG–13.3%, CRF02_AG–8.9%, and URF_B/G recombinant viruses–17.8%) using a convenience set of plasma samples taken from IDUs, living in the Lisbon metropolitan area, showed overall agreement with the previously published subtyping results for the HIV-1 circulating among the Portuguese population.^{18–21} These results disclose a high level of different URF_B/G viruses. However, this was not unexpected given the fact that (1) subtype B and G viruses are very frequently found in Portugal and (2) some high-risk behaviors associated with IDUs (e.g., needle-sharing and unprotected sex) potentiate multiple infections and, subsequently, the generation of recombinant viruses. What was surprising was, perhaps, the absence of URF with CRF02_AG sequences in the analyzed samples. Whether this reflects a sampling bias resulting from the low circulation of CRF02_AG viruses among IDUs, a limitation arising from the small number of viral targets analyzed, or any biological feature that limits the genesis and/or viability of CRF02_AG recombinants remains to be established.

Accession Numbers

The accession numbers are *gag* (FR848963–FR849006), *pol(Pr)* (FR850059–850102), *pol(RT)* (FR850289–FR850330), *vpu* (HE583228–HE583271), *env(gp120)* (FR870327–870369), and *env(gp41)* (FR870477–870518).

Acknowledgments

We would like to thank Prof. H.-G. Kräusslich for kindly providing plasmid pBD6-15. The following reagents were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: 8E5/LAV from Dr. Thomas Folks and pNL4-3 from Dr. Malcolm Martin. The work presented here is part of a larger research project involving the genetic characterization of blood-borne viruses circulating in the IDUs of the Lisbon metropolitan area, and was partially funded by Fundação GlaxoSmithKline das Ciências de Saúde (Portugal) as well as by Fundação para a Ciência e a Tecnologia (FCT/MCTES) through Unidade de Parasitologia e Microbiologia Médicas (UPMM) funds. We would also like to express our gratitude to the staff of Centro de Atendimento de Toxicodependentes das Taipas for their kind cooperation, Teresa Venenno for technical assistance, and Cristina Branco and Vera Benavente for their help with the analysis of nucleotide sequences by bootscanning.

Author Disclosure Statement

No competing financial interests exist.

References

- Butler IF, Pandrea I, Marx PA, *et al.*: HIV genetic diversity: Biological and public health consequences. *Curr HIV Res* 2007;5:23–45.
- Plantier J-C, Leoz M, Dickerson JE, *et al.*: A new human immunodeficiency virus derived from gorillas. *Nat Med* 2009;15:871–872.
- McCutchan FE: Understanding the genetic diversity of HIV-1. *AIDS* 2000;14:S31–S44.
- Rouet F, Chaix ML, Nerrienet E, *et al.*: Impact of HIV-1 genetic diversity on plasma HIV-1 RNA quantification: Usefulness of the Agence Nationale de Recherches sur le SIDA second-generation long terminal repeat-based real-time reverse transcriptase polymerase chain reaction test. *J. Acquir Immune Defic Syndr* 2007;45:380–388.
- Taylor BS, Sobieszczyk ME, McCutchan FE, *et al.*: The challenge of HIV-1 subtype diversity. *N Engl J Med* 2008;358:1590–1602.
- McCutchan FE: Global epidemiology of HIV. *J Med Virol* 2006;78:S7–S12.
- Thomson MM and Nájera R: Molecular epidemiology of HIV-1 variants in the global AIDS pandemic: An update. *AIDS Rev* 2005;7:210–224.
- Delwart EL, Shpaer EG, Louwagie J, *et al.*: Genetic relationships determined by a DNA heteroduplex mobility assay: Analysis of HIV-1 env genes. *Science* 1993;262:1257–1261.
- Arroyo MA, Hoelscher M, Sanders-Buell E, *et al.*: HIV type 1 subtypes among blood donors in the Mbeya region of south-west Tanzania. *AIDS Res Hum Retroviruses* 2004;20:895–901.
- Arroyo MA, Hoelscher M, Sateren W, *et al.*: HIV-1 diversity and prevalence differ between urban and rural areas in the Mbeya region of Tanzania. *AIDS* 2005;19:1517–1524.
- Herbinger KH, Gerhardt M, Piyasirisilp S, *et al.*: Frequency of HIV type 1 dual infection and HIV diversity: Analysis of low- and high-risk populations in Mbeya Region, Tanzania. *AIDS Res Hum Retroviruses* 2006;22:599–606.
- 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.
- 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.
- Kijak GH, Sanders-Buell E, Wolfe ND, *et al.*: Development and application of a high-throughput HIV type 1 genotyping assay to identify CRF02_AG in West/West Central Africa. *AIDS Res Hum Retroviruses* 2004;20:521–530.
- Kijak GH, Tovanabutra S, Sanders-Buell E, *et al.*: Distinguishing molecular forms of HIV-1 in Asia with a high-throughput, fluorescent genotyping assay, MHAbce v.2. *Virology* 2007;358:178–191.
- Sarkar R, Sengupta S, Mullick R, *et al.*: Implementation of a multiregion hybridization assay to characterize HIV-1 strains detected among injecting drug users in Manipur, India. *Intervirology* 2009;52:175–178.
- Watanaveeradej V, Benenson MW, Souza MD, I, *et al.*: Molecular epidemiology of HIV type 1 in preparation for a phase III prime-boost vaccine trial in Thailand and a new approach to HIV type 1 genotyping. *AIDS Res Hum Retroviruses* 2006;22:801–807.
- Esteves A, Parreira R, Venenno T, *et al.*: Molecular epidemiology of HIV type 1 infection in Portugal: High prevalence of non-B subtypes. *AIDS Res Hum Retroviruses* 2002;18:313–325.
- Esteves A, Parreira R, Piedade J, *et al.*: Spreading of HIV-1 subtype G and envB/gagG recombinant strains among injecting drug users in Lisbon, Portugal. *AIDS Res Hum Retroviruses* 2003;19:511–517.
- Palma AC, Araujo F, Duque V, *et al.*: Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. *Infect Genet Evol* 2007;7(3):391–398.
- Parreira R, Pádua E, Piedade J, *et al.*: Genetic analysis of human immunodeficiency virus type 1 nef in Portugal: Subtyping, identification of mosaic genes, and amino acid sequence variability. *J Med Virol* 2005;77:8–16.
- Tebit DM, Zekeng L, Kaptué L, *et al.*: Construction and characterisation of a full-length infectious molecular clone from a fast replicating, X4-tropic HIV-1 CRF02_AG primary isolate. *Virology* 2003;313:645–652.
- Katoh K and Toh H: Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 2008;9:286–298.
- Tamura K, Dudley J, Nei M, *et al.*: MEGA4: Molecular Evolutionary Genetic Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24:1596–1599.
- 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 inter-subtype recombination. *J Virol* 1999;73:152–160.
- Rychlik W, Spencer WJ, and Rhoads RE: Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res* 1990;18:6409–6412.
- Sheffield VC, Cox DR, Lerman LS, *et al.*: Attachment of a 40-base-pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc Natl Acad Sci USA* 1989;86:232–236.
- Abecasis AB, Martins A, Costa I, *et al.*: Molecular epidemiological analysis of paired pol/env sequences from Portuguese HIV type 1 patients. *AIDS Res Hum Retroviruses* 2011;27:803–805.
- Domingues ACV: Caracterização do Gene *pol* de Vírus da Imunodeficiência Humana Tipo (VIH-1) circulantes na Beira, Moçambique. Masters Thesis, Universidade Nova de Lisboa, Portugal, 2006: 187.
- White PA, Li Z, Zhai X, *et al.*: Mixed viral infection identified using heteroduplex mobility analysis (HMA). *Virology* 2000;27:1382–1389.
- Manigart O, Courgnaud V, Sanou O, *et al.*: HIV-1 superinfections in a cohort of commercial sex workers in Burkina Faso as assessed by an autologous heteroduplex mobility procedure. *AIDS* 2004;18:1645–1651.

Address correspondence to:

Ferdinando Freitas
Unidade de Microbiologia Médica
Instituto de Higiene e Medicina Tropical
Rua da Junqueira 100
P-1349-008 Lisboa
Portugal

E-mail: ferdinandofreitas@gmail.com