Distinguishing molecular forms of HIV-1 in Asia with a high-throughput, fluorescent genotyping assay, MHAbce v.2

Gustavo H. Kijaka,⁎ Sodsai Tovanabutra, Eric Sanders-Buell, Veerachai Watanaveeradej, Mark S. de Souza, Kenrad E. Nelson, Vidhaya Ketsararat, Vilawan Gulgo In, Manu Wera-arpachai, Somchai Sriplienchand, Chirasak Khamboonrueng, Deborah L. Birx, Merlin L. Robb, Francine E. McCutchan

a Henry M. Jackson Foundation for the Advancement of Military Medicine/US Military HIV Research Program, 1600 East Gude Drive, Rockville, MD 20850, USA
b Royal Thai Army Component, Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand
c Henry M. Jackson Foundation for the Advancement of Military Medicine, Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand
d Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA
e Lampang Hospital, Ministry of Public Health, Thailand
f Phramongkutklao Hospital, Royal Thai Army, Thailand
g Research Institute for Health Sciences, Chiang Mai University, Thailand
h Division of Retrovirology, Walter Reed Army Institute of Research, Silver Spring, MD, USA

Received 22 June 2006; returned to author for revision 10 July 2006; accepted 28 July 2006

Abstract

High-resolution HIV-1 genotyping of large sample sets is crucial to define the evolving and dynamic epidemics in Asia. Here we present MHAbce v.2, a multi-region hybridization assay that individually discriminates subtypes B, C, CRF01_AE, and virtually all of their described recombinants, based on real-time PCR using subtype-specific TaqMan probes in 8 regions throughout the viral genome. In a validation panel (n = 70), the assay performed with a sensitivity of 95.7% and specificity of 99.8%. The assay was field-tested on samples from a retrospective MTCT cohort (n = 180; Lampang Province, Northern Thailand; 1996–1998). 177/180 of the samples were typeable, and 94.4% were typed as CRF01_AE. The remaining strains represented even proportions of subtype B and B/CRF01_AE recombinants and were confirmed by sequencing, revealing early links between the heterosexual and IDU HIV-1 epidemics in Thailand. MHAbce v.2, with an area of application including China, India, Southeast Asia, and the Pacific Rim, can be used to develop a comprehensive and detailed picture of this important component of the HIV/AIDS pandemic. Published by Elsevier Inc.

Keywords: HIV-1; Molecular epidemiology; Multi-region hybridization assay (MHA); Recombinant forms; Asia; Thailand

Introduction

Asia is home to almost half of the world’s population and harbors one in five global Human Immunodeficiency Virus type 1 (HIV-1) infections (UNAIDS, 2006). Nine genetic subtypes and many inter-subtype recombinant forms circulate in the pandemic (Kijak and McCutchan, 2005; McCutchan, 2006; Peeters and Sharp, 2000), but the genetic diversity of HIV-1 in Asia is only a subset of this global diversity (Tovanabutra et al., 2004a, 2004b). Identification of HIV-1 strains in Asia is fundamental to the selection and evaluation of candidate vaccines (Thomson et al., 2002) and is important to monitor development of antiretroviral drug resistance in different subtypes (Wainberg, 2004). Molecular epidemiology has been widely applied on the continent in an effort to document the diversity of strains and their geographic distribution (reviewed in Tovanabutra et al., 2004a, 2004b).

Genotyping of HIV-1 can identify strains associated with particular risk behaviors, providing focus for intervention, and can...
reveal the social networks, often transcending national borders, in which HIV-1 spreads (Avila et al., 2002; Kalish et al., 1995). Moreover, the spread of new molecular forms, often generated through inter-subtype recombination in high-risk groups multiply exposed to HIV-1, can be specifically addressed to limit the growing complexity of the Asian epidemic (Delgado et al., 2002; Liitsola et al., 1998; Piyasirisilp et al., 2000; Rodenburg et al., 2001; Takebe et al., 2003; Tovanabutra et al., 2003). Finally, the design and evaluation of candidate vaccines will be facilitated by a complete description of the circulating strains.

Nucleotide sequences drawn from more than 13,000 HIV-1 strains from Asia are available, but only 1% of these sequences encompass the complete HIV-1 genome (http://www.hiv.lanl.gov). The overwhelming majority of HIV-1 sequences from Asia, partial or complete, are either subtype B, subtype C or CRF01_AE, a circulating recombinant form (CRF) combining subtypes A and E (Carr et al., 1996; Gao et al., 1996; Tovanabutra et al., 2004a, 2004b). Recombinants of these strains are also being documented in many countries (Piyasirisilp et al., 2000; Ramos et al., 2002; Rodenburg et al., 2001; Takebe et al., 2003; Tee et al., 2005; Tovanabutra et al., 2003, 2004a, 2004b, 2005), but the power to identify and track novel recombinant forms has been largely vested in the complete genome sequencing approach, at least to date (Carr et al., 1999; Salminen et al., 1995b).

The challenge for molecular epidemiology of HIV-1 in Asia is to provide enough genetic information per strain to distinguish the three principal strains from the recombinants derived from them, while providing adequate sampling of large numbers of HIV-1 positive individuals. Complete genome sequencing of a sufficient number of strains is impractical, but multi-region hybridization assays (MHAs) that use fluorescent, subtype-specific probes applied to multiple genome regions per strain have provided an adequate alternative technology in several complex, regional epidemics outside Asia (Arroyo et al., 2005; Hierholzer et al., 2002; Hoelscher et al., 2002; Kijak et al., 2004).

The objective of this study is to design, validate, and apply a high-throughput, fluorescent genotyping assay, MHAbce v.2, constructed specifically to distinguish all of the known molecular forms of HIV-1 in Asia, now numbering thirty. Such an approach could provide more accurate estimates of the proportions of strains, track rare strains that may begin to circulate widely, and identify new molecular forms as they arise. Here we report the design, validation, and first field application of MHAbce v.2 and show its potential to track the Asian HIV-1 epidemic with improved scope and clarity.

Results

Design of an HIV-1 genotyping assay for Asia

The objective of this study is to design, validate, and field test a high-throughput genotyping assay capable of distinguishing the known HIV-1 subtypes, circulating recombinant forms (CRFs) and unique recombinant forms (URFs) in Asia. Current knowledge of the genetic diversity of HIV-1 in Asia provided the basis for assay design. Published complete or virtually complete genome sequences from HIV-1 infections in China, India, Japan, Korea, Myanmar, and Thailand, together with partial sequences from these countries and from Bangladesh, Cambodia, Malaysia, Nepal, the Philippines, Singapore, and Taiwan were available for analysis (http://www.hiv.lanl.gov).

Fig. 1 indicates the region of Asia targeted for assay design and shows the subtype distribution and country of origin of HIV-1 strains in the complete genome database. The complexity of strains in the Asia epidemic is depicted in Fig. 1b. Virtually all of the HIV-1 infections characterized by complete or partial sequencing are either subtype B, subtype C, or CRF01_AE or are recombinants of these. CRFs combining subtypes B and C (CRF07_BC, CRF08_BC) (Piyasirisilp et al., 2000; Rodenburg et al., 2001) or subtype B and CRF01_AE (CRF15_01B) (Tovanabutra et al., 2003) have been identified, along with 18 unique B/CRF01_AE recombinant forms (Ramos et al., 2002; Takebe et al., 2003; Tovanabutra et al., 2003, 2004a, 2004b, 2005; Watanaveeradej et al., 2006), an apparent CRF07/CRF08 recombinant (Yang et al., 2003), two unique B/C recombinants (Takebe et al., 2003), one unique C/CRF01_AE recombinant (Watanaveeradej et al., 2003), and two B/C/CRF01_AE recombinants (Takebe et al., 2003).

Development of an MHA with genotyping capacity for Asia will require the design and validation of fluorescent probes specific for subtypes B, C, and CRF01_AE. Fig. 1c highlights the additional requirement for characterization of multiple strain regions in a single probe to meet the objective of distinguishing the many different recombinant forms. The genome location of probes will be an important consideration. For example, the vertical lines in Fig. 1c represent the locations of eight subtype-specific probe sets that, together, would develop hybridization patterns across the genome that would specifically distinguish virtually all of molecular forms of HIV-1 known to exist in Asia.

Subtype-specific probes

Development of the MHAbce v.2 requires probes designed for optimal recognition of the homologous subtype or CRF, with minimal cross reaction on heterologous subtypes/CRF. Another goal is the design of inner and outer PCR primers that can achieve universal amplification of all relevant HIV-1 subtypes/CRF. For probe and primer design, an alignment of virtually complete HIV-1 sequences from SE Asia and neighboring countries was used to identify regions of clustered within-subtype conservation and between-subtype variation in the eight regions of HIV-1 genome depicted in Fig. 1c.

Discrimination between subtypes/CRF depends on the generation of probe/target DNA duplexes that are stable at the annealing/extension temperature used for real-time PCR for the homologous subtype/CRF, but not for heterologous subtypes/CRF. The thermal stability of duplexes can be predicted, in part, from the length and base composition, but the actual effects of probe/target mismatches can be strongly dependent on sequence context and the nature and distribution of mismatches (Turner, 1996), and on electrolyte and detergent concentrations. Each candidate probe was tested empirically for its performance on homologous and heterologous target sequences.
Candidate oligonucleotide probes were initially synthesized without the fluorescent dye and quencher required for use in TaqMan real-time PCR (“cold probes”), together with complementary unlabeled oligonucleotides representing homologous and heterologous target sequences. Hybridization was performed under conditions to be used in real-time PCR and in the presence of SybrGreen, whose fluorescence intensity is enhanced when bound to double-stranded DNA. Thermal
dissociation profiles of the probe/target duplexes were then evaluated using SybrGreen fluorescence.

Many of the candidate probes provided the desired discrimination. Fig. 2a shows a CRF01_AE-specific probe that forms duplexes with melting temperatures (Tms) well above the 60 °C annealing/extension temperature with a CRF01_AE target sequence, but of much lower Tm with subtype B or C. In other cases, the subtype discrimination of probes needed to be improved. We used *ad hoc* introduction of an additional mismatched base, not represented in the target sequences, to improve discrimination (Kijak et al., 2004; Plantier et al., 2002). Fig. 2b shows an example of improved specificity. Once all 24 probes (3 subtypes × 8 genome regions) had been identified, inner and outer PCR primers were designed in areas of complete or near-complete sequence conservation closely flanking the probe sites. Primers were tested empirically for their ability to amplify all of the relevant subtypes and CRF (data not shown). The nucleotide sequences of the primers and probes for the MHAbee v.2 assay are given in Table 1.

**Assay validation**

The performance of MHAbee v.2 was evaluated on a panel of 70 samples for which the complete nucleotide sequence had been determined (Table S1), representing the subtypes, CRFs, and URFs circulating in SE Asia and neighboring countries. The target sequences were in the form of uncloned PCR products that had been used for complete genome sequencing or, in the case of subtype C from India, were amplified from complete HIV-1 genomes cloned in plasmid vectors.

First, 53 samples from the panel representing pure subtype B, C, or CRF01_AE strains were evaluated. Probes for subtype B, subtype C, and CRF01_AE performed with high sensitivity and specificity. The reactions with heterologous subtypes,
Table 1
Primer and probes used in MHAbce v.2

<table>
<thead>
<tr>
<th>Region</th>
<th>Sequence (5′–3′)</th>
<th>Primer/Probe Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Region p17</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer primers</td>
<td>Forward</td>
<td>AGACAGGAWCAGARGAACTTARATCATT</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>ACCCATGCATTYAAGGTTCTAGGG</td>
</tr>
<tr>
<td>Inner primers</td>
<td>Forward</td>
<td>CCAARGAAGCCTTAGAAARATAGAGGAAG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>TGCCCTTGKRSATYTGCCATAGGG</td>
</tr>
<tr>
<td>Probes (HXB2: 1133–1163)</td>
<td>CRF01_AE</td>
<td>FAM-CAGCGACGCAGCAGGCAAAGGAGCAG-BHQ1</td>
</tr>
<tr>
<td></td>
<td>Subtype B</td>
<td>FAM-CAGCGACGCAGCAGCAGCATAAGGAACAR-BHQ1</td>
</tr>
<tr>
<td></td>
<td>Subtype C</td>
<td>FAM-CAGCGACGCAGCAGCAGCATAACGGAAG-BHQ1</td>
</tr>
<tr>
<td><strong>Region pro</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer primers</td>
<td>Forward</td>
<td>AGGCCRRGGRAATTTTCYGTCAG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>TCAATTTGTGTTCTCAATCG</td>
</tr>
<tr>
<td>Inner primers</td>
<td>Forward</td>
<td>CAGAGCAGACGAGAGCCACCACG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>TTCCTGGCAATYNTCTCTATAC</td>
</tr>
<tr>
<td>Probes (HXB2: 2268–2298)</td>
<td>CRF01_AE</td>
<td>FAM-TGCGCAACGCTCCCCCCTTCAGGTTAAGATAG-BHQ1</td>
</tr>
<tr>
<td></td>
<td>Subtype B</td>
<td>FAM-TGCGCAACGCTCCCCCCTTCAGGTTAAGATAG-BHQ1</td>
</tr>
<tr>
<td></td>
<td>Subtype C</td>
<td>FAM-TGCGCAACGCTCCCCCCTTCAGGTTAAGATAG-BHQ1</td>
</tr>
<tr>
<td><strong>Region rt</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer primers</td>
<td>Forward</td>
<td>TTYTTGGGAAGTTCAATTAGGTAATCC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>AAAGTCATCCATGTAGTAGTATACAYATKTCAG</td>
</tr>
<tr>
<td>Inner primers</td>
<td>Forward</td>
<td>GGGAGATCGCATTTCAGTTTCCTT</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>YYYCTAAARGGCTCTAGAAATTCAG</td>
</tr>
<tr>
<td>Probes (HXB2: 2968–2999)</td>
<td>CRF01_AE</td>
<td>FAM-CAGGAAACGCTATTCAGGATTAG-TACGCAATACAGAG-BHQ1</td>
</tr>
<tr>
<td></td>
<td>Subtype B</td>
<td>FAM-CAGGAAACGCTATTCAGGATTAG-TACGCAATACAGAG-BHQ1</td>
</tr>
<tr>
<td></td>
<td>Subtype C</td>
<td>FAM-CAGGAAACGCTATTCAGGATTAG-TACGCAATACAGAG-BHQ1</td>
</tr>
<tr>
<td><strong>Region int</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer primers</td>
<td>Forward</td>
<td>AAAATAGCAGGAAAGATGGCCAG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CTGCTCTAAAGRTGTAYTCACGTG</td>
</tr>
<tr>
<td>Inner primers</td>
<td>Forward</td>
<td>GTGAARGMCAGCTGCTGGT</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CTACTCCYTGACTTTGGAATTGTA</td>
</tr>
<tr>
<td>Probes (HXB2: 4622–4645)</td>
<td>CRF01_AE</td>
<td>FAM-AATTCCTGTGCTGGTCTGCCAC-BHQ1</td>
</tr>
<tr>
<td></td>
<td>Subtype B</td>
<td>FAM-AATTCCTGTGCTGGTCTGCCAC-BHQ1</td>
</tr>
<tr>
<td></td>
<td>Subtype C</td>
<td>FAM-AATTCCTGTGCTGGTCTGCCAC-BHQ1</td>
</tr>
<tr>
<td><strong>Region tat</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer primers</td>
<td>Forward</td>
<td>TGCAAAACTCCTGTTTTCATTTTC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>TACTATRGTCCACACAAACTATTGCATSGA</td>
</tr>
<tr>
<td>Inner primers</td>
<td>Forward</td>
<td>GAATTGGGTGTCRCACTAGCAGAATAGG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>TGYCCCGCTTCTCCTGCCAT</td>
</tr>
<tr>
<td>Probes (HXB2: 5918–5946)</td>
<td>CRF01_AE</td>
<td>FAM-TGTTGCTGGTCCAGTAATGTTCTCCT-BHQ1</td>
</tr>
<tr>
<td></td>
<td>Subtype B</td>
<td>FAM-TGTTGCTGGTCCAGTAATGTTCTCCT-BHQ1</td>
</tr>
<tr>
<td></td>
<td>Subtype C</td>
<td>FAM-TGTTGCTGGTCCAGTAATGTTCTCCT-BHQ1</td>
</tr>
<tr>
<td><strong>Region gp120</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer primers</td>
<td>Forward</td>
<td>GTRGTATCAACTCAAYTRCTGTATAATGGAAG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>YRATGAGGAGRGCAATCATG</td>
</tr>
</tbody>
</table>

numbering 848 (2 heterologous probes × 8 genome regions × 53 samples), were uniformly negative, for a specificity of 100%. Among 424 reactions with homologous probes, 409 hybridized, for a sensitivity of 96.5% (data not shown). Probe reactivity for each subtype and region is shown in Supplemental Fig. 1.

The remaining 17 strains in the panel represented CRF15_01B, CRF08_BC, or URFs. Fig. 3 shows their detailed structure, and the patterns of probe hybridization observed. In 126 of 136 reactions, the probes hybridized correctly to their homologous subtype/CRF. Eight reactions were negative, all due to lack of probe reactivity rather than a failure to amplify, 1 reaction showed dual probe reactivity, and 1 showed only reactivity with a heterologous probe. The samples exhibiting dual probe reactivity or heterologous reactivity were URFs, and it is possible that these derive from individuals infected with more than one HIV-1 strain.

In summary, the assessment of the 70-sample panel, involving 560 homologous reactions and 1120 heterologous reactions, showed that the MHAbce v.2 performed with a sensitivity of 95.7%, a specificity of 99.8%, a positive predictive value of 99.6%, and a negative predictive value of 97.9%.

HIV-1 subtypes in Lampang Province, Thailand

In the late 1980s, the near-simultaneous introduction of CRF01_AE in heterosexually acquired HIV-1 infections and subtype B among intravenous drug users (IDUs) in Thailand set up a subtype separation by risk group that persisted for some time (McCutchan et al., 1992; Ou et al., 1992, 1993; Weniger et al., 1991, 1994). Eventually, these barriers broke down and CRF01_AE, subtype B, and their recombinants began to circulate in both IDUs and heterosexuals (Kalish et al., 1995; Tovanabutra et al., 2003). It is unclear at what stage in the epidemic the intermixing of strains became significant in the heterosexual risk group. We selected a retrospective sample set from 1996 to 1998 for analysis to determine what fraction of non-CRF01_AE strains could be detected. 180 maternal plasma samples from a study of mother-to-child HIV-1 transmission (MTCT) in Lampang Province, Thailand, were evaluated. Viral RNA extracted from plasma was used as template for the MHAbce v.2 assay, following a reverse transcription step (see Materials and methods).

In this retrospective sample set, an average of 7 out of 8 probes hybridized per sample. Fig. 4a shows the subtype distribution of strains, based on an arbitrary minimum of 4 probes positive to determine a genotype, a criterion that was attained for 177/180 (98.3%) samples (see also Supplemental Fig. 2). For 167/177 typed strains (94.4%), all of the probes that hybridized detected only CRF01_AE. There were 10 strains (5.6%) for which at least one subtype B or C probe hybridized. Among the non-CRF01_AE strains, four were classified as subtype B. Four were B/CRF01_AE recombinants, one of which had a probe hybridization pattern consistent with CRF15_01B. One sample showed dual probe reactivity in one of the genome regions, and one sample hybridized with subtype B, C, and CRF01_AE probes in different genome regions. Among the 167 CRF01_AE strains, the mean number of genome regions typed was 7.1, whereas among the non-CRF01_AE strains, the mean probe reactivity was 6.6 regions (data not shown).

To further characterize the 10 strains that were not CRF01_AE, complete genome sequencing was performed. Phylogenetic analysis of virtually complete genome sequences
Fig. 4b showed five subtype B strains, three strains clustering with CRF01_AE, one CRF15_01B strain, and one outlier strain.

Three of the subtype B strains clustered significantly with other subtype B strains from Asia (96TH_M041, 97TH_M081, and 98TH_M145), while two (98TH_M140, and 98TH_M149) fell outside this cluster, like other subtype B strains from North America or Western Europe (Kalish et al., 1994). Further analysis of the complete genome sequence by bootscanning showed that the five subtype B strains were non-recombinant (data not shown). By MHAbce v.2, four of the subtype B strains, including the three in the Asian cluster (96TH_M041, 97TH_M081, and 98TH_M145) and one of the Western Bs
a) Map showing Lampang Province with a pie chart indicating CRF01_AE n=157, Not typed n=3, and Non-CRF01_AE n=10.

b) Phylogenetic tree depicting Subtype B, Subtype D, Subtype C, and Subtype A with various branches and mutations.

c) MHAAbce typing:
- Subtype B: 96TH_M041, 97TH_M081, 98TH_M145, 98TH_M149
- CRF15_01B: 98TH_M169
- B/CRF01_AE URFs: 96TH_M005, 96TH_M043, 98TH_M171
- Other patterns: 98TH_M140, 97TH_M114

Legend for the tree branch positions and mutation points.
Strain 96TH_M043 had small subtype B regions in pol and nef. The outlier strain (98TH_M171) contained more equal proportions of subtype B and CRF01_AE, with four breakpoints across the genome. Fig. 4c shows that the probe hybridization patterns were mostly consistent with the recombinant structures derived by bootscanning.

Finally, the strain that showed probe reactivity in the pol region with both the CRF01_AE and subtype B probes (98TH_M114) yielded a pure CRF01_AE strain by complete genome sequencing. Cloning of the pol region of this sample and examination of several individual clones with MHAbce v.2 showed consistent dual probe reactivity, suggesting that the strains with dual probe reactivity constituted the majority of the viral quasispecies present. An infrequent polymorphism at the probe binding site, not typical of CRF01_AE strains in the sequence database, was responsible (data not shown).

The relationship of the newly described B/CRF01_AE recombinant structure with the prototype strains of this CRF were investigated by comparison of the location of breakpoints with MHAbce v.2. The URFs did not share any breakpoints in common with the prototype strains of this CRF (Tovanabutra et al., 2003), confirming its typing by MHAbce v.2. The URFs did not share any breakpoints in common with each other, but comparison with previously described B/CRF01_AE URFs revealed several possible shared breakpoints (Fig. 5). The most extensive network of relationships was observed for the small segment of subtype B in the first exon of tat/rev in strain 96TH_M005 from Lampang. The transition from CRF01_AE to subtype B at HXB2 position 5573 and back to CRF01_AE at HXB2 5968 is also found in strain C2254, from a community cohort in Chon Buri Province, Thailand in 2000 (Watanaveeradej et al., 2006). Other strains share one of these breakpoints, but not both. The transition from CRF01_AE to subtype B at HXB2 5573 is found in OUR740i, and from subtype B to CRF01_AE at HXB2 5968 is found in OUR033i, OUR846i and 96TH02; these four strains were found in injecting drug users in the northern Thailand Province of Chiang Mai (2001–2002) (Tovanabutra et al., 2004a, 2004b) and in Bangkok (1996) (Ramos et al., 2002). These results raise the possibility that a series of related B/CRF01_AE recombinants, with common features but different structures overall, have been circulating in Thailand since 1996.

The newly described strain 96TH_M043 from Lampang shared two of its breakpoints with R1741 from Rayong Province (Watanaveeradej et al., 2006), and one with OUR840i from Chiang Mai (Tovanabutra et al., 2004a, 2004b). Similarly, M171 has points of commonality with 98TH_MU2003SEM, from a community cohort in Bangkok, OUR740i, and 99TH_OUR2574 from Chiang Mai, and CRF15_01B (Fig. 5). It is quite interesting that 98TH_M171 and 98TH_MU2003SEM appear to share three breakpoints in common, suggesting that they may share a more recent common parental strain.

**Discussion**

Multi-region hybridization assays (MHAs) represent one solution to the growing challenge of HIV-1 genotyping in epidemics where there is co-circulation of different subtypes and the generation of many different inter-subtype recombinant strains (Arroyo et al., 2005; Hierholzer et al., 2002; Hoelscher et al., 2002; Kijak et al., 2004). These assays have been designed to bridge the gap between characterization of each HIV-1 strain by sequencing of the entire genome, on the one hand, and the requirement for adequate sampling of large and complex epidemics, on the other. A family of MHAs, each with application to a specific regional epidemic, has been developed in recent years, including MHAacd for East Africa (Arroyo et al., 2005; Hoelscher et al., 2002), MHAcrf02 for West and West central Africa (Kijak et al., 2004), MHAbf for South America (Hierholzer et al., 2002), and MHAbce v.1 for application in Asia (Watanaveeradej et al., 2006). Here we describe the redesign of the MHAbce assay (MHAbce v.2) to encompass new information about the diversity of molecular forms of HIV-1 circulating in Asia.

Elements of the newly developed MHAbce v.2 assay include use of subtype-specific fluorescent probes in eight different regions of the HIV-1 genome, re-positioning of probes to better discriminate the many inter-subtype recombinant forms that have been described, and incorporation of de-stabilizing mutations within probes, when necessary, to improve specificity. It is also important to note the high throughput of this assay. Using a 384-well plate format in an ABI7900 HT Sequence Detection System, 85 strains per week can be genotyped. This development cycle has resulted in a new, high-throughput genotyping system that performed with
sensitivity of greater than 95% and a specificity of nearly 100% on a panel of 70 strains of known genotype. Remaining challenges for assay development include evaluation of the assay on additional strains of B′, a variant of subtype B common in Asia (Graf et al., 1998; Kalish et al., 1994; Ou et al., 1992, 1993), and on additional subtype C strains since relatively few of these were used in the validation panel. The performance on subtype B strains more typical of epidemics in Western Europe and the Americas also needs to be evaluated since these may be present at low levels; in the Lampang cohort, one Western subtype B strain did not hybridize to three of the eight probes, and another was mistyped as triple recombinant.

We applied the MHAbce v.2 to an important question pertaining to the emerging inter-subtype recombinant strains in Thailand. Using a relatively low risk MTCT cohort in Northern Thailand, the degree of intermixing of CRF01_AE strains with subtype B and recombinants was evaluated. The results indicate that, by the period 1996–1998, strains other than CRF01_AE accounted for more than 5% of infections in this small study of 180 individuals. Moreover, the recombinant strains were as numerous as the subtype B strains. By comparing the structures of recombinants with others previously described in the region, we were able to establish the presence of a CRF15_01B strain in Lampang Province as early as 1998; this CRF was first identified in specimens dated from 1999 to 2002 sampled in Chiang Mai, Rayong, and Bangkok, but had apparently already circulated to some extent earlier in the North (Tovanabutra et al., 2003).

In a similar vein, three unique recombinant strains identified in Lampang in this study shared some of their breakpoints with other recombinant strains, described later in other regions of the country (Ramos et al., 2002; Tovanabutra et al., 2003, 2004a, 2004b; Watanaveeradej et al., 2006). Most of these recombinant strains were found in individuals with injecting drug use as their main risk for HIV-1 infection, but some were from individuals with heterosexual risk as well. The interconnection of IDU and heterosexual networks in Thailand is reinforced by these data. The identification of groups of recombinant strains that share some of their breakpoints in common suggests that recombinant strains may undergo further rounds of recombination, either with
the parental strains or with other recombinants. High-risk groups who are repeatedly exposed to HIV-1 and may become dually infected can be an important source of these interrelated recombinant forms. Since none of the women that we studied in Lampang reported injecting drug use (data not shown), it remains possible that the B/CRF01_AE recombinant strains that we found were transmissions from male partners with connection to IDU networks.

The wider application of the MHAbce v.2 can help to characterize an important and highly complex epidemic of HIV-1 infection in Asia and may help shed light on the social conditions that foster the emergence of new recombinant strains. Preparation for vaccine trials, which can include genotyping of HIV-1 strains in large cohorts of potential trial volunteers, is another important application for this approach. The broad geographic area of application of the assay, including China, India, Southeast Asia, and many countries of the Pacific Rim, can foster the development of a comprehensive and detailed picture generated with a common and standardized genotyping approach of an important component of the HIV/AIDS pandemic.

Materials and methods

Sequence alignment for assay development

Published full-genome HIV-1 nucleotide sequences from Asia were retrieved from the Los Alamos National Laboratory HIV-1 database (http://hiv.lanl.gov). An alignment was built using ClustalX (Thompson et al., 1997) and was manually edited using Genetic Data Environment (Smith et al., 1994).

Estimation of hybridization strength of oligonucleotide duplexes

Candidate strain-specific probes were selected using visual inspection of nucleotide alignments, and with Primer Express software v2.0 (Applied Biosystems). The thermal stability of duplexes between candidate probes and target sequences was measured by thermal dissociation curves of duplexes formed between corresponding unlabeled oligonucleotides (Sigma Genosys, St. Louis, MO) in the presence of SybrGreen (Applied Biosystems), which fluoresces while bound to double-stranded DNA. For each genomic region, at least one CRF01_AE-, one subtype B-, and one subtype C-specific probe were tested. Additional candidate probes, containing mutations introduced ad hoc that would increase the specificity by destabilizing the heteroduplexes, were also tested. For genomic regions where limited, though consistent, within-clade sequence variation was observed in probe-binding sites, additional targets that contained the observed polymorphisms were included. For each unlabeled probe/target pair, a reaction mix containing 2× TaqMan SybrGreen PCR Master Mix (Applied Biosystems), 1 μM probe, 1 μM target, and dH2O was prepared in duplicate. Dissociation curves were obtained using the following routine in a 384-well ABI PRISM 7900HT Sequence Detection System (Applied Biosystems): 95 °C for 15 s, 30 °C for 15 s, and 95 °C for 15 s, at a 10% ramp rate. The melting temperature (Tm) of the duplexes was estimated as the maximum of the peak in the −dRn/dT vs. temperature graph.

MHAbce v.2

The principle of the MHAbce v.2 is to amplify multiple short fragments throughout the HIV-1 genome and assess the hybridization of clade-specific fluorescent probes in a real-time PCR (Hoelscher et al., 2002). The amplified regions were positioned in areas of the genome that would tend to maximize discrimination among the different strains of interest (Fig. 1c). Seven regions of the HIV-1 genome were amplified in separate first-round PCRs, or RT-PCRs, if the starting material is proviral DNA or viral RNA, respectively. The evaluated regions lay within p17, protease (pro), reverse transcriptase (rt), integrase (int), first exon of tat (tat), C3 in gp120 (gp120), and gp41/nef (Fig. 1c). The first-round PCR mix contained GeneAmp 10× PCR Gold Buffer (Applied Biosystems, Foster City, CA), 200 μM each dNTP, 1.5 mM MgCl2, 400 nM of each outer primer, 1.25 U of AmpliTaq Gold DNA Polymerase (Applied Biosystems, Foster City, CA), and 5 μl of extracted DNA, in a final volume of 25 μl. The thermocycle routine was: hold at 95 °C for 10 min, then 30 cycles of 95 °C 10 s, 55 °C 30 s, and 72 °C 1 min, and a final extension at 72 °C for 7 min, in an ABI 9600 thermocycler (Applied Biosystems, Foster City, CA). The RT-PCR mix contained QIAGEN OneStep RT-PCR 5× Buffer (Qiagen, Valencia, CA), 200 μM each dNTP, 400 nM of each outer primer, 4 U of RNasin Ribonuclease Inhibitor (Promega, Madison, WI), 0.5 μl of QIAGEN OneStep RT-PCR Enzyme Mix (Qiagen, Valencia, CA), and 5 μl of extracted RNA, in a final volume of 25 μl. The thermocycle routine was: hold at 50 °C for 30 min, then hold at 95 °C for 15 min, then 30 cycles of 95 °C 10 s, 55 °C 30 s, and 72 °C 1 min, and a final extension at 72 °C for 10 min, in an ABI 9600 thermocycler (Applied Biosystems, Foster City, CA).

An aliquot of each amplicon was then distributed into one (in the cases of the p17, pro, rt, int, tat, and gp120 regions) or two (in the case of gp41/nef) separate sets of second-round PCRs. Second-round PCR was performed in a TaqMan real-time format, with universal primers, and fluorescent probes specific for subtypes B, C, or CRF01_AE. Probes for the different subtypes were used in separate real-time PCRs. The gp120 probes were 5′-end labeled with TET and 3′-end labeled with Black Hole Quencher (BHQ1). All other probes were 5′-end labeled with FAM and 3′-end labeled with BHQ1 (Biosource, Camarillo, CA). Second-round real-time PCR mixtures contained TaqMan 2× Universal PCR Master Mix (Applied Biosystems, Foster City, CA), 400 nM of each inner primer, 250 nM of probe, and 2.5 μl of first-round PCR product, in a final volume of 12.5 μl. Real-time PCR amplification was performed in a 384-well ABI PRISM 7900HT Sequence Detection System (Applied Biosystem) with the following routine: hold at 95 °C for 10 min, 40 cycles of 95 °C 15 s, and 60 °C 1 min. In the case of the gp120 region, the extension step was performed at 60 °C for 2 min. 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 \((C_F)\) was lower than 35 and if an exponential increase in the normalized fluorescence intensity for over five consecutive cycles was observed.

To assess if negative real-time PCR results were due to lack of sample amplification or lack of probe hybridization, a parallel real-time PCR containing \(2 \times \) SybrGreen PCR Master Mix, and similar amounts of inner primers and template, was performed in parallel to determine whether a PCR product had been generated. The identity of the amplicons was verified by conducting melting curves \((95 \degree C 15 \, s, 30 \degree C 15 \, s, \text{and} 95 \degree C 15 \, s, \text{at a} 10\% \text{ ramp rate})\), which distinguished the expected PCR products from primer–dimers of lower thermal stability.

**Assay validation**

A panel of 70 samples, for which a virtually full-genome sequence had been obtained, was used for assay validation (Table S1). The panel included subtypes B \((n=4)\), C \((n=4)\), CRF01_AE \((n=45)\), CRF08_BC \((n=1)\), CRF15_01B \((n=4)\), B/CRF01_AE URFs \((n=11)\), and C/CRF01_AE URFs \((n=1)\). Samples were collected mainly in Thailand, but specimens from India and China were also used. Nearly full-genome PCR amplicons were prepared as previously described (Carr et al., 1999; Salminen et al., 1995b) and were used for assay validation after dilution to copy numbers consistent with the range found in primary clinical samples. In the case of the subtype C isolates, nearly full-genome amplicons cloned into pCR2.1 vector were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH from Dr. Kavita Lole, Dr. Robert Bollinger, and Dr. Stuart Ray. Total DNA from PBMCs obtained from HIV-1 seronegative donors, and buffer blanks were included as negative controls (phosphate buffered saline).

**Field test of the assay**

A retrospective sample set was used to field test MHAuce v.2. Between 1996 and 1998, volunteers were enrolled in a prospective perinatal cohort study, aimed at evaluating the roles of HIV-1 viral burden, and of mucosal and systemic immunity, on MTCT, in Lampang Province, Northern Thailand. The study was conducted jointly by the Lampang Hospital (Ministry of Public Health, Thailand), the Research Institute of Health Sciences (Chiang Mai University, Thailand), Armed Forces Research Institute of Medical Research (Royal Thai Army component and US Army component, Bangkok, Thailand), Johns Hopkins University (Baltimore, MD, USA), Walter Reed Army Institute of Research (Rockville, MD, USA), and the Henry M. Jackson Foundation for the Advancement of Military Medicine (Rockville, MD, USA). Plasma samples collected from 180/181 participating HIV-1 infected pregnant women with an estimated gestational age of at least 15 weeks were available for subtyping with MHAuce v2. The median age was 24 years (inter-quartile interval: 22–27 years), and the median viral load was 4.32 log copies/ml (inter-quartile interval: 3.70–4.77 log copies/ml) (Roche AMPLICOR HIV-1 MONITOR Test version 1.5, Roche Diagnostics, Indianapolis, IN). The majority of the participants reported heterosexual sexual preference (179/180), and none reported previous or current injection drug use. Seventy three out of 180 women were under AZT treatment during pregnancy. Viral RNA was extracted using the MagnaPure robotic nucleic acid extraction procedure (Roche Diagnostics, Indianapolis, IN), following the manufacturer’s instructions, and processed according the abovementioned MHAuce v.2 protocol.

**Sequencing and phylogenetic analysis**

Near full-length HIV-1 amplicons were generated by nested PCR from DNA extracted from PBMCs, as previously described (Carr et al., 1999; Salminen et al., 1995b). Nucleotide sequencing was carried out with big dye terminators using an ABI 3100 capillary sequencer (Applied Biosystems), as described. Sequences were analyzed, edited, and assembled with Sequencher 3.0 (Ann Arbor, MI). Phylogenetic analysis was performed using the SEQBOOT, DNADIST, NEIGHBOR, CONSENSE modules of the Phylip software package (v3.2c), and with TREETOOL. Bootscanning (Salminen et al., 1995a) and sub-genomic neighbor joining trees were used to characterize recombinant strains. Newly described sequences were deposited in the GenBank under accession numbers DQ354112–DQ354123.

**Acknowledgments**

We are indebted to the participants and Lampang Provincial health personnel for providing clinical samples. We thank RHES and AFRIMS staff, including Dr. Arthur E. Brown, for their input in facilitating this study throughout its course. We are grateful to Dr. Rutt Chuaichoowong for his contributions to the study design and Drs. Valerie Robinson, Nopporn Sittisombut, Nakorn Dabbhasuta, Suchint Wongchoosri, and Ms. Antika Wongthanee for their valuable advice.

This work was supported through NICHD by RO1 HD34343-02 and in part by a cooperative agreement between the Henry M. Jackson Foundation for the Advancement of Military Medicine and the U.S. Department of Defense, by the National Institute for Allergy and Infectious Diseases, National Institutes of Health. The views and opinions expressed herein do not necessarily reflect those of the U.S. Army or the Department of Defense.

**Appendix A. Supplementary data**


**References**

Arroyo, M.A., Hoelscher, M., Sateren, W., Samky, E., Maboko, L., Hoffmann, O., Kijak, G., Robb, M., Birx, D.L., McCutchan, F.E., 2005. HIV-1 diversity and prevalence differ between urban and rural areas in the Mbeya region of Tanzania. AIDS 19 (14), 1517–1524.


Tovanabutra, S., Brodine, S.K., Insley, J.R., Sankale, J.L., Sanders-Buell, E.,


