Molecular Epidemiology of HIV-1 Infection and Full-Length Genomic Analysis of Circulating Recombinant Form 07_BC Strains from Injection Drug Users in Taiwan

Yu-Ting Lin,1,2,a Yu-Ching Lan,1,7,a Yen-Ju Chen,1,2 Yu-Hui Huang,1,2 Cheng-Ming Lee,1 Tze-Tze Liu,3 Wing-Wai Wong,5 Jyh-Yuan Yang,6 Chin-Tien Wang,2,4 and Yi-Ming Arthur Chen1,2

1AIDS Prevention and Research Center, 2Institute of Public Health, 3Genomics Research Center, and 4Institute of Clinical Medicine, National Yang-Ming University, 5Branch of Disease Control, Taipei City Hospital, and 6Center for Disease Control, Taipei, and 7Department of Health Risk Management, China Medical University, Taichung, Taiwan

Background. Previously, we reported that there was an outbreak of human immunodeficiency virus type 1 (HIV-1) circulating recombinant form (CRF) 07_BC among injection drug users (IDUs) in Taiwan in 2004. The objectives of the present study were to conduct a molecular epidemiological analysis and to characterize the full-length genome of the Taiwanese CRF07_BC.

Methods. Three hundred and fifty-eight patients with HIV-1/AIDS from hospitals and 133 HIV-1–infected inmates from detention centers were recruited. DNA sequencing and phylogenetic analysis were conducted to determine subtypes and evolutionary relationships. Recombination breakpoints of 2 full-length CRF07_BC strains were elucidated using a bootscanning method.

Results. Of 206 HIV-1–infected patients who received a diagnosis in 2004, 44.7% were infected with subtype B, 53.4% with CRF07_BC, and 1.5% with CRF01_AE. Ninety-eight percent (109/111) of IDUs were infected with CRF07_BC. Deletions of 7–11 amino acids in both p6gag and p6 pol proteins were noted among the Taiwanese CRF07_BC strains. The CRF_07BC strains belonged to 2 phylogenetic clusters, and the first cluster contained only CRF07_BC strains from the southern part of Taiwan.

Conclusions. The Taiwanese CRF07_BC strains had 97% full-length sequence homology with the prototype from mainland China. CRF07_BC was first introduced into the southern region in 2002 and then spread to other regions in Taiwan in 2004.

It has been estimated that >10% of the HIV-1 cases worldwide are attributable to injection drug use [1]. In parts of southwest China, Myanmar, and Manipur, HIV prevalence among injection drug users (IDUs) exceeds 70% [2]. HIV-1 infections continue to expand along heroin trafficking routes in China [3]. Before 2004, the prevalence of HIV-1 in Taiwan was considered to be low. In the preceding year, rates among first-time blood donors, military conscripts, and pregnant women were measured as 5.2, 57.0, and 12.0 cases/100,000 population, respectively [4]. Concerning the high-risk populations, data gathered in 2003 in Taiwan showed a HIV-1 prevalence rate of 0.09% for IDUs, 0.2% for female sex workers, 1.9% for patients with sexually transmitted infections, and 6.7% for men who had engaged in homosexual sex in saunas or bathhouses [4, 5]. A dramatic increase in the number of new HIV-1/AIDS cases reported to Taiwan’s Center for Disease Control (CDC) occurred in 2004; the 1521 cases represent a 77% increase from 2003 [4]. Risk factor analysis of the reported cases to Taiwan’s CDC showed that the
Table 1. Nucleotide sequences of different primers used for polymerase chain reaction (PCR) in the present study.

The table is available in its entirety in the online edition of the Journal of Infectious Diseases.

proportions of IDUs increased from 1.7% (13/773) in 2002 to 8.6% (74/861) in 2003 and to 36.4% (553/1521) in 2004 [4]. Researchers performing phylogenetic analyses of globally circulating viral strains have identified 3 distinct HIV-1 groups (M, N, and O) with 9 genetic subtypes (A–D, F–H, J, and K) within the primary M group (reviewed in [6, 7]). Furthermore, mosaic HIV-1 strains that have transformed into circulating recombinant forms (CRFs) continue to play a significant role in global and regional HIV epidemics. To date, at least 34 HIV-1 CRFs have been identified [8]. The 3 main subtypes or CRFs in Asia have been identified as B, C, and CRF01_AE [6, 9, 10]. Subtype C is most prevalent on the Indian subcontinent [9, 11, 12]. Both subtype B and CRF01_AE are dominant in Southeast Asia [13–16]. Since 1997, both CRF07_BC and CRF08_BC strains have been isolated from IDUs in several provinces of mainland China [3, 17, 18]. These 2 CRFs were presumably generated in Yunnan Province and spread northwestward to Xinjiang Province and eastward to Guangxi Province.

According to previous data collected in Taiwan, subtype B is predominant in homosexual males and IDUs, and CRF01_AE is predominant among heterosexual persons [19]. Recently, outbreaks of HIV-1 CRF07_BC infections among IDUs in Taiwan have been reported [20, 21]. However, neither study has analyzed the full-length genome of the Taiwanese CRF07_BC. Therefore, the objectives of the present study were to conduct a molecular epidemiological analysis of HIV-1 infection in Taiwan in 2004 and to characterize the full-length genome of the Taiwanese CRF07_BC. Our results indicate that >98% of the HIV-1–infected IDUs who received a diagnosis in 2004 were infected with CRF07_BC and that the Taiwanese strains share 97% nucleotide sequence homology with the prototype from mainland China. Furthermore, open reading frame sequence analysis demonstrated that all Taiwanese CRF07_BC strains had deletions of 7–11 aa in their p6<sup>env</sup> and p6<sup>pol</sup> proteins.

**SUBJECTS, MATERIALS, AND METHODS**

**Subjects.** In 2004 and 2005, the following 2 groups of HIV-1–infected persons were recruited for the present study: group A, patients attending the AIDS clinics of Taipei City Hospital; and group B, inmates from the Taipei, Nantou, and Tainan detention centers (DCs), which are located in the northern, central, and southern regions of Taiwan, respectively. The numbers of patients participating in this study were 358 for group A and 133 for group B (52 from Taipei DC, 16 from Nantou DC, and 65 from Tainan DC). Of the patients in group B, 46 (16 from Nantou DC and 30 from Tainan DC) were recruited in September 2004 and have been described elsewhere [20]; 87 (52 from Taipei DC and 35 from Tainan DC) were recruited in December 2004. During each of the DC visits, all the inmates with newly diagnosed HIV-1 infections were invited to participate in the present study, and the participation rate was 100%. Sociodemographic data, types of illegal drugs used, and risk behaviors were assessed through a self-administered questionnaire. This study was reviewed and approved by the institutional review board (ethical committee) of our university. Informed consent was obtained from all the participants. Peripheral blood mononuclear cells (PBMCs) were collected for CD4 cell counts and HIV-1 subtyping.

**HIV-1 subtyping and phylogenetic analysis.** The proviral nucleotide sequences of the gag, pol, and env regions were determined using polymerase chain reaction (PCR) with Taq polymerase (TAKARA Bio) [22, 23]. Primer pairs used are listed in table 1. DNA was extracted from PBMCs by use of QIAmp blood-extraction kits (QIAGen). The reaction conditions are listed in table 2. DNA sequencing was performed using a DNA analyzer (ABI 3700; Applied Biosystems). HIV-1 subtypes were determined on the basis of phylogenetic analyses of the gag, pol, or env sequence. Sequence alignments with different subtype reference sequences from the Los Alamos HIV database (available at: http://www.hiv.lanl.gov) were conducted using the BioEdit program (version 3.7). Phylogenetic trees were constructed using the neighbor-joining method based on Kimura’s 2-parameter distance matrix with 1000 bootstrap replicates, using the MEGA (version 3.0) and PHYLIP (version 3.5) software packages [24–26]. We used MEGA for construction of neighbor-joining trees and PHYLIP for parsimony and maximum-likelihood methods to verify the topology of taxa shown in the trees. In this study, all trees are shown as the neighbor-joining tree from MEGA.

**Full-length genomic sequencing of the CRF07_BC strains.** Initially, 4 CRF07_BC strains from Taiwanese IDUs (TW_D3 and TW_D4 from the northern region, TW_D60 from the central region, and TW_D118 from the southern region) were selected for full-length sequencing. Eventually, we obtained near-full-length sequences for the TW_D3 and TW_D60 strains and incomplete sequences for the TW_D4 and TW_D118 strains. The primer pairs for the first PCR [27] and 13 nested PCRs (table 1) were designed to cover the full-length genome.
of HIV-1 CRF07_BC. The reaction conditions are listed in table 2.

**Analysis of intersubtypic mosaicism using the Simplot program.** The near-full-length sequences of Taiwanese CRF07_BC strain (equivalent to nucleotide positions 796-9411 of the HXB2 strain) were aligned with 10 HIV-1 reference strains by use of the BioEdit program. The sequences were corrected manually to ensure that gaps did not alter the open reading frames. Phylogenetic trees were constructed as described above. Simplot software (version 2.5) was used to identify the diversity plot and recombination breakpoints [27]. The genetic distance between TW_D3 and each subtype was calculated using the Simplot software with a sliding window of 200 bp moving in increments of 20 bp along the alignment.

**Analysis of amino acid signature pattern and open reading frames of Taiwanese CRF07_BC strains.** The relationship between the Taiwanese CRF07_BC strains and their prototype was elucidated by analyzing their amino acid signature patterns. The deduced amino acid sequences from the Taiwanese CRF07_BC strains were aligned with the protein sequences from several strains belonging to the following subtypes or CRFs: subtype B, subtype C, CRF07_BC, and CRF08_BC. Because the unique recombinant forms (URFs) HH069 and HH086 were found in IDUs in eastern Yunnan Province of China [28], they were also included in the analysis. The amino acid sequences of CRF07_BC-CN54 were used as the reference for comparison.

**Statistical analysis.** The Pearson $\chi^2$ test or Fisher’s exact test was performed in univariate analysis to find the statistical significance for all comparisons between patient groups with different risk factors or between patients infected with different HIV-1 subtypes. Student’s $t$ test was used to compare the mean ages and mean CD4 cell counts between outpatients and IDUs.

**Nucleotide sequence accession numbers.** The GenBank accession numbers for the near-full-length genomic sequences of CRF07_BC strains TW_D3, TW_D60, TW_D4, and TW_D118 were DQ230841, DQ230842, EF078278, and EF078279, respectively. The GenBank accession numbers for the pol sequences were EF077820–EF077943 and EF078127–EF078277. The GenBank accession numbers for the gag sequences were EF078115–EF078126. The GenBank accession numbers for the env sequences were EF077944–EF078076 and EF078077–EF078114.

**RESULTS**

In total, 491 patients with HIV-1/AIDS from Taipei City Hospital and 3 detention centers located in the northern, central, and southern regions of Taiwan participated in this study. The vast majority of the patients with HIV-1/AIDS from the hospital were homosexual (282/354 [79.7%]); in contrast, almost all of the infected inmates were heterosexual (129/131 [98.5%]). Although 131 (98.5%) of the 133 HIV-1–infected inmates had a history of injection drug use, only 5 (1.4%) of the 358 outpatients were IDUs. A larger percentage of inmates than outpatients received a new diagnosis of being HIV-1 positive in 2004—83.8% versus 27.6%. Of the outpatients, 261 were receiving antiretroviral therapy (ART) and showed higher CD4 cell counts; only 4 inmates were receiving ART (table 3).

The HIV-1 subtypes and CRFs for the 479 patients were determined using phylogenetic analysis with gag (figure 1A), pol (data not shown), or env sequences (figure 1B). The majority of the male outpatients were infected with subtype B (96.2%), whereas 50% of the female outpatients were infected with CRF01_AE. In contrast, of the 123 male inmates, 115 (93.5%) were infected with CRF07_BC, and 8 (6.5%) were infected with subtype B; the 10 female inmates were all infected with CRF07_BC.

**Table 3. Demographic data on the patients with HIV-1/AIDS who participated in the present study.**

<table>
<thead>
<tr>
<th>Subtype/CRF</th>
<th>Heterosexual</th>
<th>Homosexual</th>
<th>Bisexual</th>
<th>IDU</th>
<th>Female*</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>10 (90.9)</td>
<td>72 (96.0)</td>
<td>7 (100)</td>
<td>2 (1.9)</td>
<td>1 (10)</td>
<td>92 (44.7)</td>
</tr>
<tr>
<td>CRF01_AE</td>
<td>0 (0)</td>
<td>2 (2.7)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (10)</td>
<td>3 (1.5)</td>
</tr>
<tr>
<td>CRF07_BC</td>
<td>1 (9.1)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>101 (98.1)</td>
<td>8 (80)</td>
<td>110 (53.4)</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>0 (0)</td>
<td>1 (1.3)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>Total</td>
<td>11 (100)</td>
<td>75 (100)</td>
<td>7 (100)</td>
<td>103 (100)</td>
<td>10 (100)</td>
<td>206 (100)</td>
</tr>
</tbody>
</table>

*Includes 1 heterosexual female infected with subtype B, 1 homosexual female infected with CRF01_AE, and 8 IDUs infected with CRF07_BC.
Figure 1. Phylogenetic analyses of HIV-1 strains from Taiwanese injection drug users. A, Consensus neighbor-joining tree obtained from 1000 bootstrap replicates of aligned gag sequences from different HIV-1 strains, corresponding to nucleotide positions 1174–1640 of HXB2. B, Consensus neighbor-joining tree obtained from 1000 bootstrap replicates of aligned env sequences from different HIV-1 strains, corresponding to nucleotide positions 7077–7657 of HXB2. Each Taiwanese (TW) circulating recombinant form (CRF) 07_BC strain is labeled with different symbols and characters to denote the patients’ characteristics, including the year of diagnosis, from 2002 (02) to 2004 (04); the sex of the patient (M, male; F, female); and the location of the detention center (N, north; C, central; S, south). The scale bars indicate the no. of nucleotide substitutions per site.
Figure 1. (Continued.)
Figure 2.  A, Phylogenetic analysis of the full-length sequences of Taiwanese circulating recombinant form (CRF) 07_BC strains. The bootstrap values for nodes a, b, and c are 100. The scale bar indicates the no. of nucleotide substitutions per site. B, Diversity plot of CRF07_BC TW_D3. The plot was based on background alignments using reference sequences derived from selected virus strains representing the most relevant HIV-1 subtypes. The standard representatives are marked by different colors, as indicated in the key. The X-axis indicates the nucleotide positions along the alignment; the Y-axis indicates the similarity of TW_D3 to the listed reference subtypes. C, Bootscanning plots of the prototypic (97CN54, top) and Taiwanese CRF07_BC (TW_D3, bottom) strains. The bootstrap values are based on 1000 replicates, using the neighbor-joining method. The X-axis indicates the nucleotide positions in the multiple alignment of the near-full-length HIV-1 sequences; the Y-axis indicates the bootstrap value (percentage). D, Schematic structure of the mosaic genome of CRF07_BC TW_D3. CA, capsid; IN, integrase; MA, matrix; NC, nucleocapsid; PR, protease; RT, reverse transcriptase.
Molecular epidemiology of HIV-1 subtypes in Taiwan in 2004. As shown in table 4, of the 206 patients who received a diagnosis of HIV-1 infection in 2004, 92 (44.7%) were infected with HIV-1 subtype B, 3 (1.5%) with CRF01_AE, and 110 (53.4%) with CRF07_BC. When we analyzed the subtypes according to risk factor, 98.1% (101/103) of the male IDUs and all 8 female IDUs were found to be infected with CRF07_BC. Only 2 IDUs were infected with subtype B. In contrast, 90.9% (10/11) of the heterosexual males, 96% (72/75) of the homosexual males, and all 7 bisexual males were infected with subtype B.

Origin and dissemination paths of Taiwanese CRF07_BC strains. As shown in figure 1, phylogenetic analyses based on either the gag or env region revealed that the Taiwanese CRF07_BC strains were clustered with CRF07_BC strains originating from mainland China. In addition, the Taiwanese CRF07_BC strains were clustered in 2 groups, with a bootstrap value of 76 in the tree constructed on the basis of the env sequences (figure 1B). As shown in figure 1B, group 1 contains CRF07_BC strains exclusively from the inmates in the detention center located in the southern part of Taiwan, whereas group 2 contains CRF07_BC strains only from inmates in the detention centers located in the central and northern regions of Taiwan.

Phylogenetic and bootscan analyses of the full-length Taiwanese CRF07_BC strains. Four CRF07_BC strains (TW_D3 and TW_D4 from the northern region, TW_D60 from the central region, and TW_D118 from the southern region) were selected for further sequence analysis. Eventually, we obtained near-full-length sequences for TW_D3 and TW_D60, and they shared 97% nucleotide sequence homology with the prototypic CRF07_BC strain (97CN54, 97CN001, 98CN009, and CNGL179) from mainland China. As shown in figure 2A, a phylogenetic tree of TW_D3 and TW_D60 strains was con-
ducted using the neighbor-joining program, and the results showed that both TW_D3 and TW_D60 strains clustered with other CRF07_BC strains from mainland China, with a bootstrap value of 100. Subsequently, the Simplot program was used to compare the nucleotide distances between the CRF07_BC TW_D3 strain and other HIV-1 subtypes, including subtypes A, B, C, D, F, G, H, J, and K. As shown in figure 2B, the diversity plot of TW_D3 showed alternating regions having high sequence homology with subtype B’ (RL42) and subtype C (95IN21068). The majority of the nucleotide sequences of the TW_D3 strain showed the highest homology to subtype C, with 5 inserted segments of subtype B’ (RL42) and subtype C (95IN21068). The recombination breakpoints were elucidated by informative site analysis, and the results showed that there were 10 recombination breakpoints in both TW_D3 and TW_D60 and that they were almost identical to those of the CRF07_BC prototype (97CN54) from China. A schematic representation of the mosaic genome of the Taiwanese CRF07_BC TW_D3 is given in figure 2D.

Analysis of the amino acid signature pattern of the Taiwanese CRF07_BC strains. To elucidate the amino acid signature pattern of the Taiwanese CRF07_BC strains, reference strains from the following different subtypes were used for comparison: subtype B, subtype C, CRF07_BC, CRF08_BC, and URFs. Compared with the prototypic CRF07_BC-CN54, there were 28 aa differences in the envelope glycoprotein gp120 in the Taiwanese CRF07_BC strains. Of them, 16 were found in the variable regions (V1, V2, V4, and V5), and the other 12 were found in the conserved regions (figure 3A). The V3 loops in the gp120 of the Taiwanese CRF07_BCs had 100% sequence homology with the Chinese CRF07_BC-CN54 strain. Amino acid substitutions in gp41 were identified at 16 different positions (16/345 [4.64%]), and 10 were located in the endodomain (figure 3A).

Regarding the Gag protein, the amino acid sequences of matrix (p17), capsid (p24), and nucleocapsid (p7) were almost identical to those in the prototypic CRF07_BC, with amino acid substitutions of 1.5% (2/132), 0.4% (1/231), and 2.8% (2/71), respectively. Two proteins were encoded in the HIV-1 p6 region: p6env (52 aa) and transframe p6pol (48 aa). It is worthy to note that there were 7–11 aa deletions in the p6env and p6pol proteins of those 4 Taiwanese CRF07_BC strains (figure 3B and 3C).

In terms of the pol gene, which encodes the protease (PR), reverse transcriptase (RT), and integrase (IN) enzymes, the amino acid substitution rates were 2.02% (2/99) for PR, 0.68% (3/440) for RT, and 1.74% (5/288) for IN. No drug resistance–related mutations were found among the 4 sequences analyzed.

**DISCUSSION**

In the present study, the molecular epidemiological data showed distinct differences in the demographic characteristics and HIV-1 subtypes between the outpatients and incarcerated populations in Taiwan. Furthermore, the CRF07_BC strain was identified as the primary circulating virus among Taiwanese IDUs in 2004. Analysis of the demographic data revealed that a higher proportion of inmates than outpatients received a new diagnosis of HIV-1 infection in 2004 (83.7% vs. 27.6%). The HIV-1 cases in outpatients accumulated over a period of time, whereas most of the HIV-infected inmates in the detention centers received their diagnoses immediately after they entered the facilities. We compared the mean CD4 cell counts between IDUs from the detention centers and from the hospital and found that there was no difference between these 2 groups (482.3 vs. 495.2 cells/μL; P = .86, Student’s *t* test).

According to the Los Alamos HIV database, there are at least 34 different HIV-1 CRFs, and most of them (23/27 [85%]) have recombination breakpoints in the gag-pol region [10]. Previously, we and another group used both the pol and env genes for the phylogenetic analysis of CRF07_BC strains in Taiwan [20, 21]. In the present study, in addition to the pol and

---

**Table 5.** Comparison of the amino acids of the V3 loop of Taiwanese circulating recombinant form (CRF) 07_BC strains with the consensus V3 sequences from CCR5- and CXCR4-using HIV-1 subtype C isolates proposed by Coetzer et al. [29].

<table>
<thead>
<tr>
<th>HIV strain</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taiwanese CRF07_BC</td>
<td>C</td>
<td>T</td>
<td>R</td>
<td>P</td>
<td>G</td>
<td>N</td>
<td>N</td>
<td>T</td>
<td>R</td>
<td>K</td>
<td>S</td>
<td>I</td>
<td>R</td>
<td>–</td>
<td>–</td>
<td>I</td>
<td>G</td>
</tr>
<tr>
<td>CCR5 usage</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>N</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>CXCR4 usage</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

**NOTE.** The amino acids within the CXCR4-using consensus that differ from the CCR5-using consensus are highlighted by X, and the crown motif is shaded. Variable (X) and identical (.) amino acids are indicated.
env genes, we used a fragment of the gag gene containing recombinational breakpoints of CRF07_BC for subtyping. The results showed that the Taiwanese CRF07_BC strains clustered with other CRF07_BC strains from China, with a significant bootstrap value of 99 (figure 1A). Therefore, the region that we used (nucleotide positions 1174–1640 of the HXB2 genome) here was suitable for the identification of CRF07_BC strains. Furthermore, we analyzed 4 near-full-length genomic sequences of Taiwanese CRF07_BC strains, and both the phylogenetic tree and bootscanning analyses confirmed that they were closely related with the CRF07_BC strains from mainland China (figure 2).

The first case of HIV-1 infection among drug users in Taiwan was reported in 1988 [4]. Before 1998, subtype B was found to be the predominant subtype among Taiwanese IDUs, although its prevalence rate was relatively low [19]. In 2004, CRF07_BC surpassed subtype B and became the predominant HIV-1 infection among IDUs in Taiwan. CRF07_BC was first detected among IDUs in Xinjiang Province of mainland China in 1997 [3, 17]. In the present study, phylogenetic analysis using either a gag or env gene fragment showed that the Taiwanese CRF07_BC strains were clustered with CRF07_BC strains isolated from Xinjiang (97CN54 [AX149647], 97CN001 [AF286226], and 98CN009 [AF286230]) and Guangxi (CNGL-179 [AF503396]) Provinces of mainland China. In addition, the phylogenetic tree constructed using the env gene showed that the Taiwanese CRF07_BC strains clustered into 2 groups, with a bootstrap value of 76 (figure 1B). Group 1 contains viruses exclusively from the southern part of Taiwan, whereas group 2 contains only viruses from the central and northern parts of Taiwan. Because group 1 contains several cases that were diagnosed in 2002 and 2003 and all of the group 2 cases were diagnosed in 2004, we speculate that CRF07_BC was first introduced in the southern part of Taiwan in 2002 and then spread further to the central and northern parts of Taiwan in 2004. We analyzed the nucleotide sequence variations of a fragment of the env gene (nucleotide positions 7077–7657 of HXB2) of the Taiwanese CRF07_BC strains and found that the intragroup sequence variations for groups 1 and 2 were 1.44% (n = 15) and 0.59% (n = 10), respectively; this finding also suggests that the group 2 viruses. Further studies with more CRF07_BC strains from Taiwan and other provinces of mainland China (including Guangdong Province) will be helpful to elucidate the origin and dissemination routes of the CRF07_BC strains in Taiwan.

Concerning the amino acid sequence variation of the Taiwanese CRF07BC strains, their V3 loops of the gp120 harbor a sequence identical to that of the Chinese CRF07_BC strains. As shown in table 5, the V3 loops of the Taiwanese CRF07_BC strains contain consensus CCR5 tropic–related amino acid residues of subtype C [29, 30]. Regarding the gag and pol coding sequences, it has been found that a −1 ribosomal frameshift event occurs at a frequency of ~5% during gag translation, resulting in Pol being translated as a Gag-Pol fusion protein [31]. Within the Gag-Pol, the p6 domain is truncated and replaced by a transframe domain referred to as p6pol [32]. It has been proposed that the p6pol protein contains a PTAP domain at its N-terminal region that binds the Tsg101 protein [33] and residues 32–46 at its C-terminal region that are sufficient for AIP1 binding [34]. Tsg101 and AIP1 proteins link HIV-1 p6 gag to the endosomal sorting complex, which is essential for viral budding. In terms of p6pol, previous reports have suggested that p6pol acts as a regulator of protease activation and that the C-terminal SFNF motif is critical for protease autocleavage [35]. In the present study, we found that the PTAP domain was conserved among all the Taiwanese CRF07_BC strains, whereas they had 7–11 aa deletions in the p6pol protein. As shown in figure 3B, the deletions were found mainly in the AIP1-binding domain of the p6pol protein. These deletions were also observed in 3 of 8 CRF07_BC isolates from Kunming, Yunnan Province, China [36]. Recently, we sequenced another 14 CRF07_BC strains from IDUs who received HIV diagnoses in 2005 and 2006 and found that all of the strains had similar deletions in the p6pol protein. In contrast, according to sequence analysis data on 127 HIV-1 strains (subtypes A–D, F–H, J, K, CRF01_AE, and CRF02_AG), only 3 subtype A strains had 2–5 aa deletions in the AIP1-binding domain [37]. Further studies are needed to elucidate the effect of these deletions on the viral fitness and transmission efficiency of the Taiwanese CRF07_BC.
Figure 3. A, Signature patterns of the envelope glycoproteins (gp120 and gp41) of circulating recombinant form (CRF) 07_BC strains in Taiwan. B and C, Alignments of the amino acid sequences of the p6 gag (panel B) and p6 pol (panel C) proteins of different CRF07_BC strains from Taiwan and mainland China. The amino acid sequence of CRF07_BC 97CN54 was used as a reference. The nos. above each panel show amino acid positions relative to the HXB2 sequence; periods (.) denote identity with the reference sequence; dashes (–) indicate gaps inserted into the sequence to maintain sequence alignment; and asterisks (*) above the nos. indicate amino acid insertions in front of the position. The amino acid residues of the signature pattern of CRF07_BC strains from Taiwan are shaded.
Acknowledgments

We thank Miss Shu-Fen Lai, Miss Shu-Yuan Liang, and Miss Kuan-Hsuan Chen of the AIDS Prevention and Research Center of National Yang-Ming University, for their technical assistance, and Prof. Donald E. Morrisky and Mr. Jon Lindemann, for their help in editing the manuscript. We also express our gratitude for the technical support of the Genome Research Center of National Yang-Ming University.

References