Sequence Analysis of the *gag-pol* Gene of Human Immunodeficiency Virus Type 1 of Intersubtype (B'/C) Recombinant Strain in Beijing, China

Jing-rong Ye,^{1,2} Ruo-lei Xin,² Li-shi Bai,² Hong-yan Lu,² Shuang-qing Yu,¹ and Yi Zeng¹

Abstract

Little is known about the molecular and biological properties of HIV-1 intersubtype B'/C in Beijing. To fill the gap, we sequenced and analyzed the *gag-pol* genes from 39 HIV-1 B'/C recombinant infectors in Beijing, China during 2007. The results show that 36 CRF07_BC and 2 CRF08_BC isolates have a structural profile identical or nearly identical to CRF07_BC or CRF08_BC according to sequences in the *gag-pol* regions. The CRF07_BC circulating in injecting drug users (IDUs) and heterosexuals forms a diverse phylogenetic tree and most isolates from homosexuals cluster together. However, all the B'/C recombinant strains were remarkable for their low interpatient diversity in *gag-pol* genes (3.1, 3.0, and 2.2% for isolates from IDUs, heterosexuals, and homosexuals, respectively). We identified I7V, E91G, N242T, and K361R in the *gag* gene and R290I (HXB2 positions) in the *pol* gene as signature amino acid substitutions characteristic of HIV-1 CRF07_BC from the Beijing lineage. In addition, one new B'/C recombinant was detected. These results may contribute to an understanding of HIV-1 in Beijing.

CRF07_BC AND CRF08_BC ARE TWO types of B'/C recombinant strains that were first documented among Chinese injecting drug users (IDUs) in 2000.¹ CRF07_BC (prototype strains, 97CN54 and 97CN001) was distributed among IDUs in Xinjiang province in northwest China,^{2,3} whereas CRF08_BC (prototype strain, 97CNGX6F) was circulating widely among IDUs in Guangxi province in southeast China.⁴ Afterward these two types of recombinant strains spread from IDUs to their sexual partners and to the general population all around the country.^{5,6} Beijing, located in the north of China, is the capital of China and has a population of 15.6 million (2010 midyear estimate). From 1985, when the first cases in Beijing were reported,⁷ to October 2008, more than 5635 cases of HIV-1/AIDS have been documented. The main mode of HIV-1 transmission in this city is reported to be by sexual contact (35.1%) and by IDUs (34.4%). Studies of the molecular epidemiology of HIV-1 in Beijing conducted in 2006 revealed that CRF07_BC was the predominant subtype followed by CRF01_AE, constituting 31.7% of all subtypes. Among IDUs and heterosexuals, CRF07_BC was also the predominant subtype, accounting for 78.9% and 38.9% of all subtypes.^{8,9}

With the rapid increase of B'/C recombinant virus infections in Beijing, taking proper measures against this pandemic was an urgent public health priority. It is also important and necessary to learn more about the genetic characterization of

current frequently transmitted HIV-1 B'/C recombinant strains. There were about 12 published B'/C recombinant *gag-pol* sequences in the Los Alamos HIV database (<http://hiv-web.lanl.gov/>), but most of them were all isolated from IDUs in Xinjiang and Guangxi province.^{10,11} Due to the extensive diversity and rapid divergence in HIV-1, new B'/C recombinant sequences must be evaluated to determine the current extent of variation and recombination in viral genomes. This was also essential for the design of an effective HIV-1/AIDS vaccine. Our laboratory is responsible for the routine diagnosis and confirmation of HIV-1 infection in the locality. To determine the genetic characteristics of HIV-1 B'/C recombinant strains and the possibility that B'/C intersubtype recombinants are present in Beijing, we amplified *gag-pol* sequences from 39 individuals infected with the HIV-1 intersubtype B'/C recombinant through different high-risk activities throughout 2007 in this city. This study may provide information relevant to vaccine development in Beijing and may assist public health workers in preventing the spread of HIV-1. The study was approved by both the study participants, who provided informed consent after being assuring of anonymity, and the Committee on Human Research at Beijing Center for Disease Prevention and Control in China.

All the blood samples used in this study were collected from HIV-1 B'/C recombinant infectors from different risk

¹National Institute for Viral Disease Control and Prevention, China Center for Disease Prevention and Control, Beijing, China.

²Beijing Center for Disease Prevention and Control, Beijing, China.

groups who lived in Beijing, China during the national molecular epidemiology survey of 2007. All plasma samples were stored at -80°C . The epidemiological characteristics of 39 B'/C recombinant HIV-1 infectors are summarized in Table 1.

The 39 samples analyzed were from 33 men, aged 21–64 years (average: 32.9 years), and 6 women, aged 24–47 years (average: 33.5 years). The 39 HIV-infected persons were from 12 provinces including Beijing, Hebei, Sichuan, Xinjiang, Inner Mongolia, Jiangsu, Jiangxi, Shandong, Hubei, Chongqing, Guangxi, and Hainan province. The isolates from Beijing, Hebei, Sichuan, and Xinjiang constitute 20.5% (8/39), 7.7% (3/39), 17.9% (7/39), and 33.3% (13/39), respectively. In addition to the Han ethnic group, which accounted for the majority of our study participants (21/39; 53.8%), HIV-1 was detected in three other minority populations including Uyghur (33.3%), Yi (10.3%), and Dai (2.6%). Of the 39 subjects studied, 17 acquired their infections through sexual contact (10 heterosexuals and 7 homosexuals) and 22 from IDUs. Among them, five patients were symptomatic and 34 were asymptomatic. All the samples analyzed were from drug-naïve patients and viral RNA was isolated from plasma by the QIAamp viral extraction kit (QIAGEN, Germany).

To determine the presence of recombination breakpoints, a 2.6-kb *gag-pol* sequence was amplified by reverse transcriptase polymerase chain reaction (RT-PCR; TaKaRa One Step RNA PCR kit) with outer primer pairs Pg-3 (5'-AGGGACTTGAAAGCGAAAG-3', 648–666 nt, HXB2) and Pg-4 (5'-

TTCCAGGATAAATCTGACTTGC-3', 3348–3369 nt, HXB2), followed by nested PCR (Takara Ex Taq PCR kit) with inner primer pairs Gag-F2 (5'-ATGGGTGCGAGAGCGTCARTATTAA-3', 790–814 nt, HXB2) and LR62 (5'-TTCTGTATGTCATTGACAGTCCAGCT-3', 3300–3325 nt, HXB2).

The amplification was carried out in a thermal cycler (170-9703, Bio-Rad) for first-round RT-PCR (50°C for 30 min and 94°C for 2 min), followed by 1 cycle (94°C for 4 min, 52°C for 2 min, 72°C for 5 min) and 35 cycles (94°C for 30 s, 52°C for 45 s, 72°C for 5 min), and a final extension for 7 min at 72°C . From the first-round PCR products, 5 μl was a template and was used for the nested PCR with inner primers and the same cycling condition as the first-round PCR was used except that the cycles were 50°C for 30 min.

The second-round PCR products were purified using spin columns (QIAGEN, Germany) and were sequenced in both directions with primers Gag-E2 (5'-TCCAACAGCCCTTTTCCTAGG-3', 2032–2011 nt, HXB2), c-gag (5'-TAGTTCCTGCTATRTCACTTCC-3', 1507–1486 nt, HXB2), pro-c3 (5'-TTGAGTGTATATGGATTTTCAGG-3', 2727–2703 nt, HXB2), pro-c1 (5'-AAAAGGGCTGTTGGAAATGTGG-3', 2018–2039 nt, HXB2), and RT-4 (5'-GGATGGCCCCAAAAGTTAAAC-3', 2597–2616 nt, HXB2) using an ABI 3730 sequencer. The sequences were aligned with previously reported HIV-1 strains of various subtypes from the Los Alamos database using Vector NTI 8.0 software (Invitrogen, USA). Multiple alignments were performed automatically by BioEdit with minor manual adjustments. Gaps were introduced to create the alignment. All the nucleotide sequences obtained were screened by HIV-BLAST (http://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html) to search for sequences in the databases and rule out potential laboratory errors.

The GenBank accession numbers of nucleotide sequences reported in this article are HQ007312–007350. The Kimura two-parameter method was used for the determination of the evolutionary distance. The reliability of the branching patterns was assessed by bootstrap analysis with 1000 replicates. Phylogenetic tree analysis was performed using the neighbor-joining method implemented by MEGA 4 software and intrapatient and interpatient distances were also calculated.¹²

Thirty-six isolates form a diverse cluster within clade CRF07_BC, with a bootstrap value of 99% and a mean interpatient genetic distance of 3.7% (range, 1.9–4.1%). Two isolates grouped with the CRF08_BC, with a bootstrap value of 99% and an interpatient genetic distance of 3.2% (range, 2.2–3.7%). One sample (07CNBJ168) isolates with cluster CRF08_BC and subtype C apart from the CRF07_BC with bootstrap values of 97% and 91%, respectively, in the *gag-pol* phylogenetic tree. This isolate was somewhat distinct from the reference subtype CRF07_BC and CRF08_BC sequences; we speculate that this isolate was a B'/C unique recombinant forms (URF) and would discuss it in the latter part of the paper.

The calculated average intragroup distance of the CRF07_BC clade was 4.3% on the DNA level, indicating that the epidemic in this area is still very young. Most of CRF07_BC HIV-1 strains from Beijing are divergent and did not cluster tightly together in the phylogenetic tree except for six homosexuals, three heterosexuals, and one IDU who were thought to have epidemiological linkages from three subclusters (Fig. 1). These results also indicate that the introduc-

TABLE 1. EPIDEMIOLOGICAL CHARACTERISTICS OF 39 B'/C RECOMBINANT STRAIN HIV-1 INFECTORS IN BEIJING

Characteristics	Cases (%)
Risk factors	
Injecting drug user	22 (56.4)
Heterosexual contact	10 (25.6)
Homosexual contact	7 (17.9)
Ethnicities	
Han	21 (53.8)
Uyghur	13 (33.3)
Yi	4 (10.3)
Dai	1 (2.6)
Ages (years)	
<20	0 (0)
20–39	33 (84.6)
40–59	5 (12.8)
>60	1 (2.6)
Sex	
Male	33 (84.6)
Female	6 (15.4)
Location (province)	
Beijing	8 (20.5)
Hebei	3 (7.7)
Sichuan	7 (17.9)
Xinjiang	13 (33.3)
Others ^a	8 (20.5)
Subtype	
CRF07_BC	33 (84.6)
CRF08_BC	2 (5.1)
URF	1 (2.6)

^aIncludes Inner Mongolia, Jiangsu, Jiangxi, Shandong, Hubei, Chongqing, Guangxi, and Hainan province.

FIG. 1. Phylogenetic analysis of HIV-1 *gag-pol* gene sequences. The phylogenetic tree was constructed using neighbor-joining methods (MEGA 4) for the *gag-pol* region. Samples of heterosexual contact, homosexual contact, and IDUs are shown by black solid triangles, circles, and diamonds in the tree. The black solid squares indicate reference sequences from the Los Alamos HIV sequence database. Bootstrap values above 70% are shown at the respective nodes.

tion of these viruses into Beijing was not likely to be the result of a single event. As far as ethnicity was concerned, no distinct segregation of CRF07_BC viruses was found among any of the ethnic groups. The interpatient genetic distances of the *gag-pol* genes among IDU isolates were also compared with HIV-1 sequences from patients with heterosexual and homosexual transmission. However, all the B'/C recombinant strains were remarkable for their low interpatient diversity in the *gag-pol* genes (3.1, 3.0, and 2.2%, respectively). The interpatient diversities within *gag*, p1, p2, p6, p7, p17, p24, *pol*, RT, and *pro* were calculated and were 3.3%, 1.9%, 6.5%, 4.4%,

3.7%, 4.9%, 2.3%, 2.8%, 2.7%, and 2.5%, respectively, and p2, p6, and p17 were the highly variable *gag* domain.

To identify the presence of recombination breakpoints in samples, bootscanning analysis was performed by SimPlot (version 3.2) using a window size of 200 nucleotides, a step size of 20, and the Kimura (two-parameter) distance model. Putative parental strains of different subtypes were selected from the Los Alamos HIV database (<http://www.hiv.lanl.gov/content/index>). The background subtype sequences in this analysis were RL42 (B'), C.95IN21068 (subtype C), and H.BE.93.VI991 (subtype H). Bootscanning analysis showed

FIG. 2. RIP analysis of the *gag-pol* coding region of 07CNBJ168. RIP analysis (window size, 300; threshold for statistical significance, 90%; gap handling, STRIP) was based on background alignments using reference sequences CRF01_AE(CM240), subtype A (Q23_17), subtype B' (RL42), subtype C (95IN21), and subtype F1 (MP411). The standard representatives are marked by different colors as indicated. The x-axis indicates the nucleotide positions along the alignment. The y-axis indicates the similarity of 07CNBJ168 to the listed reference subtypes. Color images available online at www.liebertonline.com/aid.

that breakpoints located in the gag-pol region of 36 CRF07_BC and 2 CRF08_BC isolates were found at almost identical positions within CRF07_BC and CRF08_BC according to isolates from IDUs living in Xinjiang and Guangxi provinces. This is in agreement with previous studies conducted in 2004.¹³

Phylogenetic analysis revealed that one isolate (07CNBJ168) clustered with CRF08_BC and subtype C apart from CRF07_BC with bootstrap values of 97% and 91%, respectively, in the gag-pol phylogenetic tree. This isolate was somewhat distinct from the reference CRF07_BC and CRF08_BC sequences, reflected by its position between subtype C and CRF08_BC in the phylogenetic tree, suggesting probably new forms of B'/C recombination (Fig. 1). The gag-pol gene sequence of 07CNBJ168 was analyzed using the Recombinant Identification Program (RIP) available at the HIV sequence database. During RIP analysis, the query sequence was aligned to a set of gag-pol reference sequences of all HIV-1

subtypes and a comparison of the query sequence was done to obtain evidence of the viral subtype and intersubtype recombination. RIP analysis showed that the virus is the product of recombination between the subtype B' and C virus (Fig. 2). Bootscanning analysis using SimPlot software showed that there were three B'/C intersubtype recombination sites at the gag-pol region and the subtype C portions were closest to the Indian isolate 95IN21068, and the subtype B' portions were closest to isolate RL42 from Yunnan, China (Fig. 3). The positions of the recombination sites of 07CNBJ168 in the gag-pol region were approximately at nucleotide positions 1187, 1331, and 2659 of reference strain HIV-1 HXB2. The B'/C recombinant (07CNBJ07168) contained two segments from subtype B' inserted into the backbone of subtype C.

The phylogenetic trees constructed from subregions defined by these breakpoints confirmed the shifted subtype between subtypes B' and C in gag-pol (bootstrap values = 96%

FIG. 3. Bootscan and phylogenetic analyses of 07CNBJ168 sequences. The bootscan analysis was done using subtype B' and C reference sequences and one subtype H sequence as an outgroup. For phylogenetic analysis, the sequence of each recombinant virus was aligned with reference sequences from the Los Alamos HIV database. The alignment was segmented according to breakpoints identified for each of the recombinant viruses. Neighbor-joining trees of each segment are presented. The sequence of recombinant virus is marked with a black solid circle in each tree. The number of informative sites of the query sequence matching the subtype B' and C sequence for each region is shown in the corresponding color in a separate block. The HXB2 coordinates for each segment of recombinant virus are tabulated. Color images available online at www.liebertonline.com/aid.

and 96%, respectively). Phylogenetic analysis revealed that each segment of recombinant virus was found to cluster with corresponding segments of subtype B' or C viruses in agreement with the subtype assigned by bootscan analysis (Fig. 3).

However, within the CRF07_BC cluster there were three subclusters designated as subcluster I, II, and III with bootstrap values of 99%, respectively, suggesting possible epidemiological linkage. There are six members in cluster I (07CNBJ182, 312, 492, 511, 523, and 550), two members (07CNBJ548 and 587) in cluster II, and two members (08CNBJ040 and 07CNBJ489) in cluster III. Within subclusters II and III, 07CNBJ548 was a 32-year-old female from Jiangxi province, 07CNBJ 587 was a 34-year-old male from Beijing, 08CNBJ040 was a 47-year-old female from Hebei province, and 07CNBJ 489 was a 33-year-old male from Shandong province. Five members of subcluster I were reported to have acquired the infection through homosexual contact. The nucleotide sequences of amplified fragments were further translated into amino acid sequences for comparison. Amino acid signature sequences were identified using VESPA.¹⁴

We assessed whether CRF07_BC sequences from Beijing had amino acid substitutions characteristic of their origin. We found that the CRF07_BC Beijing sequence differed at five amino acids from that of other CRF07_BC sequences. Based on an abundance of at least 70%, we identified I7V, E91G, N242T, and K361R in the *gag* gene and R290 I (HXB2 positions) in the *pol* gene as signature amino acid substitutions characteristic of the Beijing lineage. Two deletions of two and five amino acids were also identified at 7 and 90 (HXB2 positions) in the *gag* gene. There are 21 subcluster-specific amino acid alterations in subcluster I members compared to CRF07_BC reference sequences. These alterations in the *gag* and *pol* gene are I7V, K62N, T90A, G126R, M228I, N242T, A248T, S252G, K361R, T469I, L486I, S498L, S499L, and E43K, S53A, F54N, E183K, K187E, R290I, E403V, and K404Q (HXB2 positions), respectively. This unique subset of subcluster-specific amino acid alterations formed a specific signature for subcluster I. All these discrete amino acid variation patterns were seen in the *gag-pol* gene sequences from Beijing, suggesting that these strains have a plausible common ancestor or are subjected to specific selection pressure.

The first case of HIV-1 infection of China was reported in 1985 in Beijing, the capital of China, which lies in the north of the country far from Xinjiang and Guangxi provinces where the B'/C recombinant predominated for over 12 years. Previous studies of the molecular epidemiology of HIV-1 in Beijing revealed that the predominant subtype of HIV-1 infection was CRF07_BC, followed by CRF01_AE and subtype B, which account for 31.7%, 30.4%, and 21.7%, respectively.^{8,9} The remaining subtypes, which included subtypes A, B', C, and CRF08_BC, constituted only about 16.2% of all samples.

In the present study, 39 nucleotide sequences of the 2.6-kb *gag-pol* gene were successfully amplified, sequenced, and phylogenetically analyzed. The results showed that 36 isolates clustered with CRF07_BC and two isolates grouped with CRF08_BC. All these 38 specimens show a structural profile identical or nearly identical to CRF07_BC or CRF08_BC according to isolates in the *gag-pol* regions. By doing phylogenetic and bootscan analyses, we identified a B'/C URF (07CNBJ168) with its *gag-pol* gene that contained two segments from subtype B' inserted into the backbone of subtype C. Isolate 07CNBJ168 was a 56-year-old male from Hubei

province who was infected with HIV-1 through heterosexual contact. We do not know where or when he was infected with HIV-1, but we speculate that he became infected about 3–5 years ago because his CD4⁺ T cell count was low (346 cell/ μ l). We think that the recombination 07CNBJ168 may not take place in Beijing because subtypes B' and C were not the most predominant HIV-1 subtype in Beijing, constituting only 11.8% and 1.9% in a molecular epidemiological study; the prerequisite of recombination of HIV-1 is that several subtypes of the virus cocirculate in a region. We therefore speculate that the URF was introduced into Beijing from other provinces of China.

Further studies with more B'/C recombinant strains from Beijing and from neighboring provinces of China and full-length sequence analysis of B'/C strains will be helpful in elucidating the origin and dissemination routes of the B'/C recombinant strains in Beijing. Yang *et al.*¹⁵ reported that diverse forms of unique recombinants between subtypes B' and C, and even the second-generation recombinants composed of CRF07_BC and CRF08_BC, have been reported among IDUs in Yunnan province. Here, we reported the findings from one B'/C URF in Beijing far from the epicenter, which was considered as Yunnan in China. The emergence of a genetically recombinant form among these individuals might serve as the basis for the rapid spread of HIV-1 in the city. With the rapid expansion of this virus in Beijing and other regions in China, we believe that our findings will have implications for understanding the new trend of the HIV-1 pandemic in Beijing.

Acknowledgments

This study was supported by China megaprojects in infectious diseases (2008ZX10001-009 and 2008ZX10001-014). We thank all laboratory members of the National Institute for Viral Disease Control and Prevention of China and the Beijing Center for Disease Prevention and Control for their support in carrying out this work.

Author Disclosure Statement

No competing financial interests exist.

References

- Shao Y, Zhao F, Yang W, *et al.*: The identification of recombinant HIV-1 strains in IDUs in southeast and northwest China. *Chin J Exp Clin Virol* 1999;13:109–112.
- Su L, Graf M, Zhang Y, *et al.*: Characterization of a virtually full length HIV-1 genome of a prevalent intersubtype (C/B') recombinant strain in China. *J Virol* 2000;74:11367–11376.
- Xin R, He X, Xing H, *et al.*: Genetic and temporal dynamics of human immunodeficiency virus type 1 CRF07_BC in Xinjiang, China. *J Gen Virol* 2009;90:1757–1761.
- Piyasirisilp S, McCutchan FE, Carr JK, *et al.*: A recent outbreak of human immunodeficiency virus type 1 infection in southern China was initiated by two highly homogeneous, geographically separated strains, circulating recombinant form AE and a novel BC recombinant. *J Virol* 2000;74:11286–11295.
- Li X-J, Kusagawa S, Xia X, *et al.*: Molecular epidemiology of the heterosexual HIV-1 transmission in Kunming, Yunnan province of China suggests origin from the local IDU epidemic. *AIDS Res Hum Retroviruses* 2005;21(11):977–980.

6. Lin H-H, Shih Y-L, Liu Y-C, *et al.*: An epidemic of HIV type I CRF07_BC infection among injection drug users in Taiwan. *J Acquir Immune Defic Syndr* 2006;42(2):248–255.
7. Zeng Y, Fan J, Zhang Q, *et al.*: Detection of antibody to LAV/HTLV-III in sera from hemophiliacs in China. *AIDS Res* 1986;2(Suppl 1):S147–149.
8. Ye JR, Xing H, Liu HL, *et al.*: Subtype and sequence analysis of *gag* and *env* genes among HIV-1 strains circulating in Beijing residents during 2006. *Chin J Epidemiol* 2007;28(6):586–588.
9. Ye JR, Xing H, Sun WD, *et al.*: Subtype and sequence analysis of *gag* genes among HIV-1 strains circulating in Beijing immigration during 2006. *Chin J Microbiol Immunol* 2007;27(10):938–941.
10. Meng Z, He X, Xing H, *et al.*: Construction and characterization of an infectious molecular clone of HIV type 1 CRF07_BC. *AIDS Res Hum Retroviruses* 2008; 24(2):259–264.
11. Meng Z, Xing H, He X, *et al.*: Genetic characterization of three newly isolated CRF07_BC near full-length genomes in China. *AIDS Res Hum Retroviruses* 2007;23:1049–1054.
12. Tamura K, Dudley J, Nei M, *et al.*: Molecular evolutionary genetic analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24:1596–1599.
13. Guan Q, Wei M, Huang H-L, *et al.*: Sequence analysis of *gag-pol* gene of HIV-1 B/C recombinant viruses in China. *Natl Med J China* 2004;84(5):387–391.
14. Gupta RM, Prasad VV, Rai A, *et al.*: Analysis of HIV type 1 subtype C full-length *gag* gene sequence from India: Novel observations and plausible implications. *AIDS Res Hum Retroviruses* 2005;21:1066–1072.
15. Yang R, Kusagawa S, Zhang C, *et al.*: Identification and characterization of a new class of human immunodeficiency virus type 1 recombinants comprised of two circulating recombinant forms, CRF07_BC and CRF08_BC, in China. *J Virol* 2003;77:685–695.

Address correspondence to:

Yi Zeng

National Institute for Viral Disease Control and Prevention

China CDC

No.100, Yingxin Street

Xuan Wu District

Beijing 100052

China

E-mail: zengy@public.bta.net.cn