Epidemiology reveals Zhaotong City as the hub of human immunodeficiency virus type 1 transmission from the Yunnan province to other regions in China

Li Ren,1,2,3† Bingbui Wang,4† Kunmei Gong,2 Pan Liu,5 Shiyi Zhou,6 Li Zhang,4 Xueshan Xia4,4* and Kunhua Wang1,6,*

Abstract

The Yunnan province in China has a high incidence of human immunodeficiency virus-1 (HIV-1) infection. Zhaotong City is located in the Yunnan province, a neglected ‘important region’. In this study, we evaluated the unique molecular epidemiological characteristics of HIV-1 infection in Zhaotong City. We collected 305 serum samples from HIV-infected patients in Zhaotong City between May 2015 and April 2016. A total of 122 samples were selected for HIV-1 gag-pol gene amplification, of which 88 were successfully amplified and sequenced for phylogenetic and phylogeographic analysis. Circulating recombinant forms 07_BC (CRF07_BC, 23 cases, 26.14 %) and CRF08_BC (49 cases, 55.68 %) were the predominant subtypes; the high proportions of these two subtypes differed from those elsewhere in the Yunnan province. The other subtypes were CRF01_AE (11 cases, 12.5 %), B (one case, 1.14 %) and unique recombinant forms (four cases, 4.55 %). Phylogeographic analysis of the CRF07_BC and CRF08_BC subtype strains revealed that Zhaotong was one of the regions in which CRF07_BC and CRF08_BC entered initially. The CRF08_BC strain originated from this region closer to the ‘root’ position of the phylogenetic tree. Thus, Zhaotong City may have been an important channel in the transmission route of HIV-1 from Yunnan to other parts of the country. Based on this unique distribution of HIV-1 subtypes in Zhaotong City, the epidemic outbreak in this area may have played an important role in the spread of CRF07_BC and CRF08_BC subtypes.

INTRODUCTION

Human immunodeficiency virus (HIV) infection is a serious threat to public health worldwide. According to data from the 2016 China Country factsheets, approximately 880000 people were infected with HIV by the end of 2016 [1]. CRF01_AE, CRF07_BC, CRF08_BC and B (B') comprise the majority of subtypes according to an extensive genotypic analysis of 31022 HIV-infected patients from China [2, 3]. Many domestic and international studies on the HIV epidemic have been performed using data on individuals in the Chinese province of Yunnan. Yunnan is situated along the drug-trafficking routes that channel heroin into China from southeast Asia’s opium-producing ‘Golden Triangle’ region [4]. Since the first HIV-1 epidemic in China was identified among intravenous drug users (IDUs) in 1989, HIV infection has spread from rural to urban areas, from border communities to those deeper within China, and from IDUs into the general population. Patterns of drug trafficking have spread unusual recombinant HIV-1 subtypes first isolated in Yunnan to distant regions of China [5]; thus, the Yunnan province is considered the epicentre of HIV-1 in China [6]. By the end of 2016, approximately 93 437 people were estimated to be living with HIV in Yunnan [7].
The molecular epidemic of HIV-1 has changed rapidly in Yunnan Province. China’s initial HIV outbreak was caused by HIV subtype B and subtype B’ strains imported from Thailand to Yunnan by drug trafficking through Myanmar [8–10], with subtype B subsequently becoming the dominant strain among IDUs in the region [11, 12]. The subtype distribution changed with the introduction of a subtype C lineage from India; this strain then became the most widely circulating one in the region [13]. Coincident with the emergence of subtype C, there was an increase in the prevalence of CRF01_AE due to the increased frequency of HIV-1 sexual transmission [14]. Co-circulation of subtypes B and C in this region led to generation of the B/C inter-subtype recombinants CRF07_BC and CRF08_BC in Yunnan, which in turn facilitated further spread into other regions during the 1990s, mainly through IDUs [15]. This transmission route ultimately caused the HIV/AIDS epidemic in China [16]. In the early 2000s, CRF08_BC and CRF01_AE were predominant in IDUs and those with sexually transmitted infections, respectively [6, 14].

Drug-trafficking activities have been implicated in the spreading of CRF07_BC and CRF08_BC across China through two different overland heroin trafficking routes: CRF07_BC [17] northwestern to the Xinjiang province, and CRF08_BC [18] eastward to the Guangxi province [5, 15]. Bayesian coalescent analyses suggested that CRF08_BC arose around 1991, which is 10 years after subtype C of Indian origin was first introduced into the Yunnan province, and that CRF07_BC evolved from CRF08_BC in the same region a few years later (ca. 1993). CRF08_BC circulation was initially limited to Yunnan before spreading in 1995 to nearby Guangxi and to Liaooning provinces. In contrast, CRF07_BC appears to have spread more rapidly to various geographically disparate regions of China [16]. Yunnan plays an important role as an entry point for heroin smuggling into China [15]. Through drug trafficking, some genotypes originally found in Yunnan have spread to neighbouring provinces, such as Guangxi (CRF08_BC) [18] and Sichuan (CRF07_BC) [19], then to Xinjiang (CRF07_BC) [17] and Henan (subtype B’) [20].

Previous reports indicated that HIV-1 infections in the Yunnan province were concentrated in the border area or within special groups, including cross-border populations [21–24] and long-distance truck drivers [25], and few reports have described HIV genotypes in the inland regions. However, the circulation of HIV-1 in Zhaotong City, which is located in the northeast region of the Yunnan province bordering Sichuan in the northwest and Guizhou in the east, has not been studied extensively since the first detection of HIV-1 in the region in 1997 [26]. Zhaotong is located along the drug transmission route and is distant from the border between China and other southeast Asian countries; this region may be the smallest area of the Yunnan province that is affected by external HIV-1 strains. Thus, HIV epidemiology in Zhaotong may thus be distinct from that in other regions of the Yunnan province, and may reflect a more primitive HIV epidemiological status. However, there have been few reports of HIV genotypes in this region.

To address this gap in our knowledge, we performed comprehensive and systematic molecular epidemiological studies of HIV-1 in Zhaotong City in order to clarify the genotype distribution and evolution characteristics of the major HIV-1 epidemic subtypes, and to provide a theoretical basis for developing AIDS prevention strategies, clinical treatments, and anti-HIV drugs and vaccines.

**RESULTS**

**Demographic characteristics of cases in this study**

From May 2015 to April 2016, a total of 305 HIV-infected patients were recruited at Zhaotong City, Yunnan province. The mean age of all patients was 42 years (42.3±14.4 years), and the male to female ratio was 1.82:1 (197/108). All HIV-infected patients had received at least 6 months of highly active antiretroviral therapy (HAART). In order to obtain more sample sequences for further analysis, we selected 122 patients with higher viral load (sampling rate, 40%) for gag-pol gene amplification.

Overall, 32 gag-pol sequences were successfully obtained from 53 HIV-positive samples in 2015, and 56 sequences were obtained from 69 samples in 2016. The 88 HIV-1-positive samples that had been successfully amplified were from nine different counties or districts of Zhaotong (Fig. 1), and samples from the remaining two counties were not obtained. The mean age of these 88 patients was 45 years (45.0±16.2), and the male to female ratio was 1.84:1 (57/31). Among the 78 cases of known transmission routes, 61 were infected through heterosexual activities, 11 were infected through intravenous drug use, five were infected through mother-to-child transmission (MCT), and one was infected through homosexual activities (Table S1, available in the online version of this article).

**Distribution of HIV-1 subtypes in Zhaotong**

Phylogenetic analysis showed that among the 88 study samples, the recombinant CRF07_BC (23 cases, 26.14%) and CRF08_BC (49 cases, 55.68%) strains originated from Yunnan and were the predominant subtypes (Fig. 2), accounting for 81.82% of total cases, with 11 cases of CRF01_AE subtype (12.50%), one case of B subtype, and four cases for which the subtype(s) could not be identified (Fig. 2).

CRF08_BC strains are currently only prevalent in local regions, including Yunnan and Guangxi. However, in this study, all CRF08_BC strains and the strains originated from Yunnan were distributed evenly in the cluster, indicating a close genetic relationship. Compared with CRF08_BC, the distribution of CRF07_BC was broader in China and included Yunnan, Xinjiang, Liaoiming, Guangdong and Guangxi. With the exception of a few strains of CRF07_BC that had a smaller genetic distance from some regions
(including Xinjiang and Liaoning), we found in this study that the majority of strains formed a cluster with the epidemic strains in Yunnan. CRF01_AE strains and those from Sichuan, Liaoning and Vietnam formed separate clusters, indicating diverse origins of local CRF01_AE strains.

In addition, recombination analysis was conducted on the gag-pol sequences of the four unclassified samples (ZT151172, ZT151159, ZT160130 and ZT160212); this showed that strains ZT151172, ZT160130 and ZT160212 were second-generation recombinant strains of CRF08_BC and CRF07_BC. The recombination pattern is shown in Fig. 2. The fourth case, ZT151159, was a recombinant strain of CRF07_BC and B subtypes (Fig. 2). Details of the recombination of these strains can be found in Supplementary Material (Fig. S1).

**Comparison of clinical data between subgroups**

Next, we compared the characteristics of the major epidemic subtypes (i.e. CRF07_BC, CRF08_BC and CRF01_AE) that were identified in this study (Table S1, available in the online Supplementary Material). As shown in Table 1, there were no differences in sex between the three subtypes ($P>0.05$). All CRF01_AE subtype strains were sexually transmitted, whereas the other two subtypes were also transmitted through IDUs and MCT. A comparison of the WHO HIV clinical stage revealed no significant difference in clinical stage among subgroups ($P=0.233$). The patients included in this study had all received more than 6 months of HAART in the hospital; thus, the viral load was low, and the number of CD4+ cells was high. There were no significant differences in CD4+ counts among the three groups. Moreover, after HAART, viral loads in the three subtypes were not significantly different.

**Phylogenetic characteristics of CRF07_BC and CRF08_BC strains**

To further elucidate the molecular characteristics of CRF07_BC and CRF08_BC strains from Zhaotong City, we carried out Bayesian phylogenetic and molecular clock analysis of their C fragments (HXB2:790–1191; 1642–2010). In the
maximum clade credibility (MCC) tree (Fig. 3), CRF07_BC, CRF08_BC and C subtypes were clustered together and distinct from other subtypes. This indicates that the two subtypes originated during the late 1980s, and that the CRF08_BC subtype originated slightly earlier. Both CRF07_BC and CRF08_BC originated from the C subtype, and the earliest epidemic strain was found in the Yunnan province.

The CRF07_BC subtype in the MCC tree was further divided into four clusters. Cluster I strains were the oldest epidemic strains, including several strains from Yunnan, Xinjiang and Taiwan. Within MCC tree clusters, strains from the Yunnan province appeared earlier, which is consistent with a previous study showing that the Yunnan province was the origin of the CRF07_BC subtype [27]. Clusters II and III were populated by rapidly spreading strains from Yunnan, Liaoning, Guangdong, Guangxi, Sichuan and Xinjiang; five strains that were identified in the current study belong to these two clusters. In contrast, cluster IV showed various regional characteristics, such as the Sichuan epidemic cluster, epidemic clusters of northern China and northeastern China (including Liaoning, Hebei and Beijing), and the Xinjiang epidemic cluster. Cluster IV strains that were identified in the current study were phylogenetically closer to those from Sichuan, Xinjiang and other regions of Yunnan, indicating the diversity of their origins.

In the existing reference sequence, the earliest CRF08_BC subtype strain was also derived from the Yunnan province, indicating that it also originated in the Yunnan province. Epidemic strains of the CRF08_BC subtype can then be subdivided into two clusters. Cluster I was mainly prevalent in the Guangxi province, while cluster II was mainly prevalent in the Yunnan province. In this study, CRF08_BC strains were evenly distributed in these two clusters. Unlike CRF07_BC, strains of the CRF08_BC subtype were more concentrated and relatively close in the MCC tree. In cluster II, the strains in this study were separated from other strains and distinct from strains that originated from other regions of Yunnan. In cluster I, strains from Zhaotong were closer to the ‘root’ position of the phylogenetic tree, suggesting that the virus may have spread from Zhaotong to Guangxi.

Fig. 2. Maximum likelihood (M-L) phylogenetic tree constructed with MEGA software using the HIV-1 gag-pol gene sequences (2.6 kb) from 88 cases. Samples used in the study are labelled with black triangles in the tree, and the suspected recombinants are marked by blue dashed frames. Recombinant identification analysis on four undefined strains (ZT151172, ZT151159, ZT160130 and ZT160212) was performed using the Map-Draw tool, which is available at the Los Alamos HIV sequence database (iphHM). The calibration bar in the bottom left corner of every tree is the genetic distance scale.
We found that both CRF07_BC and CRF08_BC originated from the Yunnan province. CRF07_BC has circulated in most parts of China, particularly in northeast and northwest areas, while CRF08_BC is more restricted to Yunnan, Guangxi and their surrounding regions. Notably, the Zhaotong strains could be observed in most epidemic clusters of the MCC tree (Fig. 3). This suggests that Zhaotong may be an important hub for these two subtypes, thereby facilitating their spread to surrounding areas.

Table 1. Comparison of demographic information and clinical data between the main subtypes in this study

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>CRF01_AE (n=11)</th>
<th>CRF07_BC (n=23)</th>
<th>CRF08_BC (n=49)</th>
<th>P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, no. (male/female)</td>
<td>6/5</td>
<td>14/9</td>
<td>33/16</td>
<td>0.684*</td>
</tr>
<tr>
<td>Age, year (mean±sd)</td>
<td>40.2±10.4</td>
<td>45.9±21.1</td>
<td>44.8±14.4</td>
<td>0.613†</td>
</tr>
<tr>
<td>Transmission category, no.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sexual</td>
<td>11</td>
<td>12</td>
<td>36</td>
<td>0.120*</td>
</tr>
<tr>
<td>IDUs</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>MCT</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Others/unknown</td>
<td>0</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>WHO stage, no.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>5</td>
<td>11</td>
<td>29</td>
<td>0.233*</td>
</tr>
<tr>
<td>II</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>1</td>
<td>7</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>CD4+, cells µl⁻¹ (mean±sd)</td>
<td>268.7±253.4</td>
<td>256.2±161.5</td>
<td>246.5±138.2</td>
<td>0.917†</td>
</tr>
<tr>
<td>HIV viral load, log₁₀ copies ml⁻¹ (mean±sd)</td>
<td>3.3±0.8</td>
<td>4.0±0.8</td>
<td>3.8±0.8</td>
<td>0.094†</td>
</tr>
</tbody>
</table>

*Chi-square statistic and Fisher’s exact tests were applied.
†One-way analysis of variance was applied.

Fig. 3. Maximum clade credibility (MCC) tree of CRF07_BC and CRF08_BC strains in the study. A time-scaled MCC tree was constructed based on the gag-pol sequences (2.6 kb) of all CRF07_BC and CRF08_BC strains from the present study together with other reference sequences from GenBank. All subtypes of HIV-1 were involved in MCC tree construction. Except CRF07_BC, CRF08_BC and subtype C, the reference strains of other subtypes were combined together, named ‘Outgroup’ in the MCC tree.
DISCUSSION

Since the first case of HIV infection was reported in Dehong Prefecture in 1989, the Yunnan province was thought to have the highest HIV prevalence of any Chinese province and was deemed the centre of the HIV epidemic in China [5]. The regions of Dehong, Honghe, Lincang, Wenshan and Dali have the most serious HIV epidemics [28]. The B* subtype, which originated in Thailand, is the earliest epidemic genetic subtype to be identified in the Yunnan province. Subsequently, the C subtype from India spread into Yunnan and became the predominant epidemic subtype [13]. B* and C recombined in Yunnan and the surrounding areas, resulting in the major recombination subtypes CRF07_BC and CRF08_BC [16]. These two recombinants are also the most prevalent within B and C subtypes in China. However, the CRF01_AE recombinant was thought to be the most prevalent subtype in many areas or distinct populations in the Yunnan province [22, 25].

Many reports of HIV in the Yunnan province have focused on the border areas of the region (for example, Dehong and Honghe) [29, 30], or special populations (such as IDUs and homosexuals) [31, 32]. However, reports of HIV infections and epidemics in the inland regions of the Yunnan province, including Zhaotong City, are scarce. Zhaotong City is located at the border of Yunnan, Guizhou and Sichuan provinces, and is an important channel through which personnel and trading goods exit the Yunnan province and enter the Sichuan province. Although the incidence of HIV infection in this region is purportedly rising, the rate is significantly lower than that of other regions in the Yunnan province. However, prior to the current study, the distribution of HIV genetic subtypes in the region was unknown.

Here we found that the CRF07_BC and CRF08_BC subtypes, which originated from Yunnan in the 1990s, were the most prevalent subtypes in the region, accounting for 81.82% of all cases. Notably, an HIV-1 epidemic with CRF07_BC and CRF08_BC as the predominant strains has not been reported elsewhere in the Yunnan province. In contrast, the CRF01_AE subtype, is the most prevalent of the other areas, accounting for only 12.50% of cases in this study. Interestingly, the C subtype, which has been reported nationwide, was not found in this study. This difference may be due to the special geographical location of Zhaotong. Moreover, CRF07_BC has become a major epidemic strain in some areas of Sichuan [33], which shares a border with Zhaotong. CRF01_AE is the main epidemic subtype predominant in the border areas of Yunnan. CRF01_AE most likely originated in southeast Asian countries, from where it entered China [34, 35]. CRF07_BC and CRF08_BC subtypes are thought to have originated in Yunnan or its surrounding areas. With the exception of large-scale epidemics of CRF07_BC in the provinces of Xinjiang and Liaoning, this subtype has been primarily identified in Yunnan and Guangxi [27]. Therefore, we suggest that the significant difference in subtype distribution in Zhaotong compared with other areas of Yunnan is due to the preservation of primitive epidemic strains in the city, which is rarely affected by HIV epidemic strains from Southeast Asia. Additionally, this finding may be related to the HIV epidemic subtypes in inland provinces.

In addition, we also found four strains that could not be classified by maximum likelihood (M-L) phylogenetic tree analysis, since the recombination patterns of their gag-pol gene sequences had not been reported previously. Thus, further studies of the full-length sequence and recombination analysis are needed for identification of these particular strains. Our comparative analysis of the clinical data of the three major epidemic subtypes in this study did not reveal significant differences between recombinants. This finding may be related to the fact that all patients received HAART.

Bayesian phylogenetic and molecular clock analyses of CRF07_BC and CRF08_BC in this study primarily analysed the origin and evolution of these two subtypes. In this study, the partial sequence of the C subtype backbone in the HIV-1 gag-pol gene fragment used for analysis and the evolutionary rate of this fragment (3.76 × 10⁻³ substitutions/site/year) were all based on previous results from Feng et al. [27]. CRF07_BC was first discovered in Yunnan and Taiwan, then in Xinjiang which were the earliest provinces this strain entered (cluster I). In a second stage, the strain spread to other provinces such as Sichuan, Liaoning, Jiangsu, Zhejiang, Guangdong, Guangxi and Hebei (clusters II and III). At the third stage, large-scale epidemics occurred in Xinjiang, Liaoning, Sichuan and Northern China (cluster IV). In this study, a small number of CRF07_BC strains was found in clusters II and III, indicating that Zhaotong was one of the first regions in which the subtype strain was found. Strains in cluster IV were closely linked to the epidemic strains in Sichuan and Xinjiang, suggesting that Zhaotong may be an important channel through which the CRF07_BC spread from Yunnan to Xinjiang and Sichuan. In the MCC tree, the CRF08_BC subtype strains could be divided into two clusters, namely those of either Guangxi or Yunnan origin. In cluster II, the strains from Zhaotong City formed a unique class, indicating differences from strains originating from other areas in Yunnan. The strains in cluster I that originated from Zhaotong were closer to the ‘root’ position of the phylogenetic tree, suggesting their possible spread from Zhaotong into Guangxi.

Here, we explored the HIV genetic subtypes in a neglected ‘important region’, i.e. Zhaotong City, in Yunnan and elucidated the phylogenetic characteristics of the main epidemic subtypes CRF07_BC and CRF08_BC. However, although we expect that our findings represent the distribution of subtypes in this region, our sample size was small. Moreover, the samples were collected from May 2015 to April 2016, and changes in HIV subtype distribution over time were not assessed. Subsequent sample collection and epidemiological analyses are still ongoing. Unfortunately, we did not find corresponding gene sequences of HIV epidemic strains in the Guizhou province, which shares a border with Zhaotong, in the HIV database. Since the patients included
in this study were all treated with antiviral therapy, sequences were not obtained from some of the samples due to a low viral load. Further studies are needed to overcome these limitations.

In summary, we have shown that CRF07_BC and CRF08_BC subtypes were the predominant epidemic subtypes in Zhaotong City within the Yunnan province. Further phylogenetic analysis revealed a potential critical role of Zhaotong in the spread of CRF07_BC and CRF08_BC subtypes from Yunnan to other provinces in China. These findings greatly increase our understanding of HIV epidemic subtypes in Zhaotong, and provide relevant data to support the prevention, control and treatment of HIV in the region. Moreover, our results also provide a basis for studies on the origins of CRF07_BC and CRF08_BC subtypes in Yunnan, as well as their spread in China.

METHODS

Sample collection
Altogether, 305 cases of HIV infection treated in the First People’s Hospital of Zhaotong City from May 2015 to April 2016 were recruited in this study. All patients were diagnosed with HIV-1 infection. The serum samples were collected from patients who had received at least 6 months of HAART. Demographic and clinical data were collected. Serum was separated from the venous blood of the patients immediately after collection in the hospital laboratory, and stored at −80 °C for further use.

HIV-1 RNA extraction, amplification and sequencing
Viral RNA was extracted from 100 µl serum using a High Pure Viral RNA Kit (Roche, Los Angeles, CA, USA) according to the manufacturer’s instructions. The RNA was subjected to reverse transcription and nested PCR to amplify the fragments of HIV-1 gag-pol using our previously reported method (2.6 kb) [29]. The obtained PCR products were authenticated by electrophoresis and purified using an agarose gel DNA extraction kit (Takara, Dalian, China) for commercial sequencing (Invitrogen, Beijing, China).

Phylogenetic and recombination analyses
Sequences were assembled and edited using the integrated Clustal X 1.83 program [36] and BioEdit Ver 7.0.8 [37]. The reference sequences derived from the HIV Sequence Database (https://www.hiv.lanl.gov/components/sequence/HIV/search/search.html) included the major HIV-1 subtypes and circulating recombinant forms, particularly the strains previously characterized in China and southeast Asian countries. Phylogenetic analyses were performed using the M-L tree method based on the Kimura 2-parameter model with 1000 bootstrap replicates in MEGA version 6.0 [38]. The HIV-1 recombinant structure was screened using the Recombinant Identification Program available from the HIV database (www.hiv.lanl.gov/content/sequence/ RIP/RIP.html). The suspected novel recombinant strains were subjected to boot scanning and informative-site analysis using SimPlot version 3.5.1 with boot-scan window sizes of 200 bases, a step size of 20 bases and 100 replicates [39].

Bayesian phylogenetic and phylogeographic analyses
To investigate the phylogenetic and evolutionary relationships of CRF07_BC and CRF08_BC strains in this study, Bayesian phylogenies for the subtype C region in gag-pol genes (HXB2: 790–1191 and 1642–2010) were inferred under the GTR+Γ4 substitution model using Markov chain Monte Carlo sampling, as implemented in BEAST v.1.7.4 with an uncorrelated log-normal relaxed molecular clock model. The subtype C region (HXB2: 790–1191 and 1642–2010) used for Bayesian phylogenetic analyses was first reported by Feng et al. [27]. A Bayesian skyline plot coalescent model with 500 million steps was used to estimate the divergence times to the most recent common ancestor of the respective subtype C and its recombinant lineages. Convergence was checked using Tracer v1.5 (http://tree.bio.ed.ac.uk/software/tracer), which ensured that all parameters had an effective sample size value of more than 200. MCC trees were visualized using the program FigTree v1.4.0 (http://beast.bio.ed.ac.uk) with the same calculation, and 83 CRF07_BC reference sequences, 51 CRF08_BC reference sequences and 12 subtype C reference sequences were downloaded from GenBank for analysis.

Statistical analysis
Statistical analysis was conducted using SPSS, version 12.0 (SPSS, Chicago, IL, USA). Characteristics were compared between the groups using chi-square tests, Fisher’s exact tests, one-way analysis of variance and nonparametric Kruskal–Wallis tests. Results with P-values of less than 0.05 were considered statistically significant.

Funding information
This study was supported by grants from the National Natural Science Foundation of China (81660094; 81460509); the Key Science and Technology Planning Project of Yunnan Provincial Science and Technology Department (2016FC005); the Joint Funds of the National Science Foundation of Yunnan Province and Kunming Medical University (2017FE467-211); 2017FE467(132); the Science and Technology Project (2016RA014). The funding organizations had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgements
The authors would like to express immense gratitude to all patients, doctors and nurses associated with this study.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
Written informed consent was obtained from all patients before sample collection. The study was approved by the institutional ethical committee of the First Affiliated Hospital of Kunming Medical University.

References
2. Ren L, Wang HW, Xu Y, Feng Y, Zhang HF et al. Sequencing of Gag/Env association with HIV genotyping resolution and HIV-