Identification of Two Similar Novel HIV-1 Recombinant Forms (CRF01_AE/CRF07_BC) among Intravenous Drug Users in Guangxi, China

Fei Zhang\textsuperscript{a}  Yao Yang\textsuperscript{a}  Bingyu Liang\textsuperscript{a, b}  Yuan Yang\textsuperscript{a, b}  Qiuyu Wei\textsuperscript{a}  Peijiang Pan\textsuperscript{a, b}  Li Ye\textsuperscript{a, b}  Hao Liang\textsuperscript{a, b}

\textsuperscript{a}Guangxi Key Laboratory of AIDS Prevention and Treatment & Guangxi Universities Key Laboratory of Prevention and Control of Highly Prevalent Disease, School of Public Health, Guangxi Medical University, Nanning, China;  
\textsuperscript{b}Guangxi Collaborative Innovation Center for Biomedicine, Life Science Institute, Guangxi Medical University, Nanning, China

Keywords
HIV-1 · Recombinant · Intravenous drug users

Abstract
New kinds of HIV-1 circulating recombinant forms (CRFs) and unique recombinant forms (URFs) earn a great prevalence in China nowadays. In this study, we identified 2 similar URFs (2016GXNNIDU037 and 2019QZLSIDU253) both isolated from intravenous drug users (IDUs) in Guangxi, China. Phylogenetic analysis of the near full-length genome (NFLG) revealed 2 URFs both clustered with CRF01_AE but setting up a monophyletic branch, supporting a high bootstrap value. Bootscan analysis and subregional recombinant analysis found that the NFLG of 2016GXNNIDU037 and 2019QZLSIDU253 were both composed of CRF01_AE and CRF07_BC, with 3 CRF07_BC mosaic segments inserted into CRF01_AE backbones. The CRF01_AE segments of the 2 URFs clustered with a previously reported cluster 2 lineage of CRF01_AE. The 5 recombinant breakpoints of the 2 URFs were quite similar. Distinct from CRF01_AE/CRF07_BC URFs reported before, 2016GXNNIDU037 and 2019QZLSIDU253 are new evidence of a high genetic variety of HIV-1 in Guangxi, which may pose new challenges to HIV-1 prevention and molecular epidemiological surveillance in China.

Introduction
HIV/AIDS has been a major public health problem threatening human health since the first patient was reported in 1981 [1]. Two types of HIV have been identified, including HIV-1 (M, N, O, and P groups) and HIV-2 (A to H groups). HIV-1 holds the predominant position as responsible for most HIV epidemics worldwide because of its primary features of high genetic variation and dynamic recombination [2]. The M group of HIV-1 has given rise to the most widespread HIV/AIDS epidemic in the world, and it is further divided into 11 subtypes (A1, A2, B, C, D, F1, F2, G, H, J, and K). In addition to subtype variation, the appearance of recombinant forms exacerbates the complexity of HIV-1 genotypes. To date, 101 circulating recombinant forms (CRFs), as well as numerous unique recombinant forms (URFs), have been reported in the Los Alamos National Laboratory HIV database (https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html), particularly in areas where multiple HIV-1 subtypes and CRFs are circulating at the same time.
The HIV-1 CRFs epidemic in mainland China is more severe than in other countries. It has been demonstrated that CRF07_BC, CRF01_AE, and CRF08_BC were the main subtypes circulating in China and were responsible for over 80.0% of total HIV-1 infections [3]. In addition, an increasing number of new CRFs have prevailed on a small scale in China, including CRF55_01B, CRF59_01B, and CRF69_01B [4–6]. The prevalence of multiple subtypes in the same area and the same high-risk population provided opportunities for HIV-1 recombination. Previous studies have shown that needle sharing by intravenous drug users (IDUs) may lead to coinfection with multiple strains of HIV-1 in the same individual [7].

Guangxi province shares a border with Vietnam and the "Golden Triangle," which is an entry point for drug trafficking routes and cross-border migration of sex workers. This provides opportunities to cultivate new HIV-1 strains and increase transmissions. Notably, Guangxi suffers the second-highest HIV-1 burden in China [8]. Previous research proved that CRF01_AE and CRF07_BC were the 2 main strains responsible for the transmission of HIV-1 in Guangxi [9]. CRF01_AE was first identified in Yunnan by several female sex workers who returned from Thailand in the early 1990s [10] and primarily circulated via IDUs, soon switching to a sexual mode of transmission by heterosexuals [11]. Similar to CRF01_AE, the origin of CRF07_BC was also traced to Yunnan from an HIV-1 subtype C sero-positive IDU’s blood sample [12] and was estimated to spread to Guangxi in 1994 [13]. Nowadays, CRF01_AE dominates the sexually transmitted HIV-1 epidemic in Guangxi, while CRF07_BC is mainly prevalent among IDUs [3, 14].

Previous studies had found 4 CRF01_AE/07_BC recombinant forms in Guangxi, isolated from 2 heterosexuals and 2 homosexuals, respectively [7, 15–17]. However, to date, no recombinant CRF01_AE/07_BC sequence has been identified from IDUs. In this study, we acquired 2 similar near full-length genome (NFLG) sequences from IDUs in Guangxi, which may be new URFs consisting of CRF01_AE and CRF07_BC.

**Materials and Methods**

In this study, 2 similar novel HIV-1 recombinant forms involving CRF01_AE and CRF07_BC were isolated from IDUs and designated as 2016GXNNIDU037 and 2019QZLSIDU253, respectively. The 2016GXNNIDU037 was from a 49-year-old male, married, of the Han ethnicity with a CD4+ T-cell count of 560 cells/µL, while the 2019QZLSIDU253 was from a 42-year-old male, unmarried, of the Han ethnicity with a CD4+ T-cell count of 551 cells/µL. Both samples were from the routine HIV-1 drug-resistance surveillance program in Guangxi, China.

In China, partial pol (RT and PR region) sequences were used for routine drug-resistance surveillance to track the HIV-1 subtypes and drug-resistance dynamics. These 2 samples were from the routine HIV-1 drug-resistance surveillance program in Guangxi, China. Initially, we downloaded reference sequences from the Los Alamos National Laboratory HIV Sequence Database (https://www.hiv.lanl.gov/components/sequence/HIV/search/search.html) to construct the maximum likelihood phylogenetic trees with all pol sequences of drug-resistance surveillance project for subtype identification. However, we found that these 2 sequences were not clustered with any subtyped reference sequences, indicating that they may be novel recombinant strains. Then, we submitted these 2 partial pol sequences to the Los Alamos National Laboratory HIV Sequence Database online tool HIV BLAST (https://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html), jumping profile Hidden Markov Model (http://jphmm.gobics.de/), and Recombinant Identification Program (https://www.hiv.lanl.gov/content/sequence/RIP/RIP.html) to identify possible recombinant. The results showed that these 2 sequences were multi-subtype recombinant strains, and their recombination patterns and breakpoints were completely different from those of the identified recombinant subtypes, suggesting that they might be a URF. We then amplified and analyzed the NFLG sequences of these 2 samples.

First, the proviral DNA was isolated from the patient’s white blood cells using the AIDLAB DNA KIT (AIDLAB, Beijing, China) according to the manufacturer’s instructions and then amplified with TaKaRa LA Taq (TaKaRa, Dalian, China) by the nested PCR in both of the 2 rounds. This included an initial denaturation at 94°C for 2 min, followed by 35 cycles at 94°C for 15 s, at 60°C for 30 s, at 68°C for 4 min 10 s, a final extension at 68°C for 10 min, and preserved at 8°C. The positive PCR products were purified and sequenced by Tianyi Huiyuan Bioscience & Technology Inc.

The chromatogram data were cleaned and assembled using Sequencher v5.4.6 (Gene Codes Corp., Ann Arbor, MI, USA). The online tool Quality Control (https://www.hiv.lanl.gov/content/sequence/QC/index.html) was first used to rule out possible cross-contamination. Then, the NFLG sequences were submitted to the online tool HIV BLAST to determine whether the same recombinant sequences had already been identified. To further confirm that 2016GXNNIDU037 and 2019QZLSIDU253 were new URFs, reference sequences (subtypes A1, A2, B, C, D, F1, F2, G, H, J, K, N, CRF01_AE, CRF07_BC, and CRF08_BC) were downloaded from the Los Alamos National Laboratory HIV Sequence Database. All nucleotide sequences were first aligned using the online tool HIV Align (https://www.hiv.lanl.gov/content/sequence/VIRALIGN/viralign.html) and then were fine-tuned manually using BioEdit and AliView v1.25 software. The NFLG phylogenetic trees were constructed using FastTree V2.1 based on the approximately maximum likelihood method with the general time-reversible method and adjusted with Figtree V1.4.3. Only branches with bootstrap values >0.7 were considered as credible.

The basic constituent of our sequence was identified by the online tool Recombinant Identification Program. Then, we performed

Identification of Two Novel HIV-1 URFs

DOI: 10.1159/000517052

Intervirology 2022;65:58–66
preliminary recombinant analysis and breakpoint determination utilizing the online tool jumping profile Hidden Markov Model, as well as conducting BootScan with Simplot V3.5.1., and then the subregion phylogenetic trees were constructed by the method mentioned above. All the phylogenetic trees were embellished with MEGA X and Adobe Photoshop Creative Suite 6. The special genomic structure map of this novel HIV-1 recombinant form was drawn with the online Reombinant HIV-1 Drawing Tool (https://www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html).

Fig. 1. The phylogenetic tree based on NFLG sequences of 2016GXNNIDU037 and 2019QZLSIDSUD253. The phylogenetic tree of NFLG sequences was constructed using the approximately maximum likelihood method with the GTR model in FastTree V2.1, as an external command of AliView. The 2016GXNNIDIU037 and 2019QZLSIDSUD253 sequences were gathered with CRF01_AE but form a distinct monophyletic branch supported by a high bootstrap value of 1. Bootstrap value >0.7 would be considered stable in the phylogenetic tree. The 2016GXNNIDU037 and 2019QZLSIDSUD253 were marked by a red solid circle. NFLG, near full-length genome; GTR, general time reversible; CRF, circulating recombinant form.
Fig. 2. Bootscan results of the novel identified CRF01_AE/CRF07_BC. a Bootscan plots of 2016GXNNIDU037 using CRF01_AE, subtype B, CRF07_BC, and subtype F1 as references. The parameters of Bootscan analysis were 400 bp on window size and 40 bp on step size.

b Bootscan plots of 2019QZLSIDU253 using CRF01_AE and CRF07_BC as putative parental reference sequences and subtype N as outgroup. The parameters of Bootscan analysis were 400 bp on window size and 20 bp on step size. CRF, circulating recombinant form.
Results

Finally, we acquired 2 NFLG sequences of 8,997 bp (from 636 to 9,600, relative to HXB2) and 9,005 bp (from 631 to 9,743, relative to HXB2) from 2016GXNNIDU037 and 2019QZLSIDU253, respectively. The NFLG phylogenetic trees show that both 2016GXNNIDU037 and 2019QZLSIDU253 clustered with CRF01_AE but forming a monophyletic branch with relatively high stability (bootstrap value = 1), which indicated that 2016GXNNIDU037 and 2019QZLSIDU253 may not belong to any existing HIV-1 subtype and may be a novel recombinant form (Fig. 1). The recombinant analysis proved that both 2016GXNNIDU037 and 2019QZLSIDU253 were composed of 6 interlaced mosaic segments, including CRF01_AE (II, IV, and VI) and CRF07_BC (I, III, and V), with 5 unique breakpoints relative to the HXB2 coordinate (Fig. 2).

Subregion phylogenetic analysis confirmed that the novel mosaic structure of 2016GXNNIDU037 is as follows: ICRF07_BC (HXB2, 790–1,840 nt), II CRF01_AE (HXB2, 1,841–2,543 nt), III CRF07_BC (HXB2, 2,544–3,084 nt), IV CRF01_AE (HXB2, 2,930–6,534 nt), V CRF07_BC (HXB2, 6,535–8,358 nt), and VI CRF01_AE (HXB2, 8,359–10,000 nt).
Identification of Two Novel HIV-1 URFs

CRF01_AE and CRF07_BC were clustered with a cluster 2 lineage of CRF01_AE/CRF07_BC. The subregions of II, IV, and VI were clustered with a cluster 2 lineage of CRF01_AE, while the subregions of I, III, and V were highly related to the CRF07_BC, respectively. The subtypes of all reference sequences were annotated in the right side of the trees. The scale bars were labeled at the bottom of each tree, and bootstrap values of necessary branches have been added. Several branches of reference sequences were compressed for a cleaner look. jpHMM, jumping profile Hidden Markov Model; GTR, general time reversible; CRF, circulating recombinant form.

8,359–9,411 nt) (Fig. 3b). It was found that 2016GXNNIDU037 and 2019QZLSIDU253 shared similar breakpoints, and the subregions of I, III, and V were highly related to the CRF07_BC, while the subregions of II, IV, and VI were clustered with a cluster 2 lineage of CRF01_AE defined by Feng et al. [18].

The breakpoints of 2016GXNNIDU037 and 2019QZLSIDU253 were completely different from those of the previously reported CRF01_AE/CRF07_BC recombinant viruses. The final map results of the special genomic structure of HIV-1 are shown in Figure 4. Three CRF07_BC mosaic segments were inserted into a CRF01_AE backbone.

Fig. 3. Subregional phylogenetic trees of the novel identified CRF01_AE/CRF07_BC. Subregional phylogenetic trees for 6 segments determined by jpHMM and Bootscan analysis were constructed with FastTree V2.1 based on the approximately maximum likelihood method with the GTR model and adjusted with Figtree V1.4.3. Each segment of 2016GXNNIDU037 and 2019QZLSIDU253 were marked by a red solid circle and a red solid triangle, respectively. The subtypes of all reference sequences were annotated in the right side of the trees. The scale bars were labeled at the bottom of each tree, and bootstrap values of necessary branches have been added. Several branches of reference sequences were compressed for a cleaner look. jpHMM, jumping profile Hidden Markov Model; GTR, general time reversible; CRF, circulating recombinant form.
in Guangxi was reported in 2016, the backbone of which was CRF07_BC and inserted with 6 CRF01_AE segments [7]. The second one was found in 2017, with a CRF01_AE backbone and 5 CRF07_BC segments inserted [16] and the latest 2 novel CRF01_AE/07_BC recombinant viruses, which both kept the CRF07_BC parental backbone with 3 CRF01_AE segments inserted, were found in 2019 [15, 17]. There has been a tendency for the CRF01_AE and CRF07_BC recombinant forms to become increasingly frequent in each risk group, which means more CRFs in the future.

IDU was prevalent among HIV/AIDS patients of Guangxi 10 years ago, which may be due to its location along a major heroin trafficking route linking Guangxi with Yunnan and Vietnam and also to its close proximity to the Golden Triangle, the world’s major heroin-producing area [20]. Regional convenience as well as the demographic complexity of drug users may promise suitable conditions for HIV transmission [21]. Meanwhile, considering the high variation in the HIV-1 genotype in Guangxi, there will be continual and recurrent emergence of novel recombinant forms among drug users and their associates, which may pose new challenges to HIV prevention and molecular epidemiological surveillance in China.

Discussion

The CRF01_AE cluster 2 is mainly prevalent among IDUs and heterosexuals in southeast and southern China as well as Northern Vietnam [18]. In this study, all CRF01_AE segments clustered within a cluster 2 lineage of CRF01_AE, which is consistent with the previous research that CRF01_AE cluster 1 and 2 dominate the epidemic of CRF01_AE in Guangxi [9]. Moreover, CRF07_BC remains the second most prevalent strain responsible for HIV-1 infections among IDUs [19], homosexual men, and heterosexuals [9]. The sequences in our study were both from IDUs, and the 2019QZLSIDU252 reported needle-sharing behavior and commercial sex before HIV-1 infection, and the 2016GXNNIDU037 admitted to having unsafe sex with a regular sexual partner who still used drugs after being diagnosed with HIV-1 infection. Multiple high-risk behaviors occurred simultaneously, resulting in the recombination between different HIV-1 strains.

With the insertion of 3 CRF07_BC segments into a CRF01_AE backbone, the 2 sequences of this study are not as complex as those in the previous studies of Guangxi. The first URF containing CRF01_AE/07_BC in Guangxi was reported in 2016, the backbone of which was CRF07_BC and inserted with 6 CRF01_AE segments [7]. The second one was found in 2017, with a CRF01_AE backbone and 5 CRF07_BC segments inserted [16] and the latest 2 novel CRF01_AE/07_BC recombinant viruses, which both kept the CRF07_BC parental backbone with 3 CRF01_AE segments inserted, were found in 2019 [15, 17]. There has been a tendency for the CRF01_AE and CRF07_BC recombinant forms to become increasingly frequent in each risk group, which means more CRFs in the future.

IDU was prevalent among HIV/AIDS patients of Guangxi 10 years ago, which may be due to its location along a major heroin trafficking route linking Guangxi with Yunnan and Vietnam and also to its close proximity to the Golden Triangle, the world’s major heroin-producing area [20]. Regional convenience as well as the demographic complexity of drug users may promise suitable conditions for HIV transmission [21]. Meanwhile, considering the high variation in the HIV-1 genotype in Guangxi, there will be continual and recurrent emergence of novel recombinant forms among drug users and their associates, which may pose new challenges to HIV prevention and molecular epidemiological surveillance in China.
In conclusion, we identified 2 similar novel recombinant virus forms of HIV-1 both isolated from male IDUs, which were distinct from the CRFs and URFs reported previously. Five similar recombinant breakpoints between the CRF01_AE and CRF07_BC reference strains were found in the 2016GXXNNIDU037 and 2019QZLSIDU253. The emergence of CRF01_AE and CRF07_BC recombinant forms might suggest high genetic variation among HIV-1 in Guangxi, warning us to continuously supervise HIV-1 molecular epidemiologic dynamics and gather enough information for vaccine design and to provide effective suggestions for accurate control.

Conflict of Interest Statement
The authors have no conflicts of interest to declare.

Funding Sources
This study was supported by the National Key Science and Technology Project of China (Grant No. 2018ZX10010002-001-006), National Natural Science Foundation of China (Grant No. 82060610), Guangxi Natural Science Foundation (Grant No. 2018GXNSFAA138070), and Guangxi Bagui Scholar (to Junjun Jiang).

Author Contributions
H.L. and Y.Y. conceived and designed the study. F.Z. and Y.Y. conducted the data analysis, literature review, and drafted the manuscript. B.L. and Y.Y. involved in data collection and interpretation. Q.W. and P.P. assisted with data management and data analysis. All the authors contributed to the revision of the manuscript and approved the final version.

Statement of Ethics
Written informed consent was signed before sample collection, and this study was reviewed and approved by the Human Research Ethics Committee of Guangxi Medical University (Ethical review No. 2013-130). The research was conducted ethically in accordance with the World Medical Association Declaration of Helsinki. The NFLG sequence of 2016GXNNIDU037 (Accession no. MN452901) and 2019QZLSIDU253 (Accession no. MN643057) have been submitted to GenBank.

References
1 Gottlieb MS, Schropp R, Schanker HM, Weisman JD, Fan PT, Wolf RA, et al. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N Engl J Med. 1981; 305(24):1425–31.
2 Li X, Li W, Zhong P, Fang K, Zhu K, Musa TH, et al. Nationwide trends in molecular epidemiology of HIV-1 in China. AIDS Res Hum Retroviruses. 2016;32(9):851–9.
3 He X, Xing H, Ruan Y, Hong K, Cheng C, Hu Y, et al. A comprehensive mapping of HIV-1 genotypes in various risk groups and regions across China based on a nationwide molecular epidemiologic survey. PLoS One. 2012; 7(10):e47289.
4 Han X, An M, Zhang W, Cai W, Chen X, Takebe Y, et al. Genome sequences of a novel HIV-1 circulating recombinant form, CRF55_01B, identified in China. Genome Announc. 2013;1(1).
5 Li X, Ning C, He X, Yang Y, Xing H, Hong K, et al. Genome sequences of a novel HIV-1 circulating recombinant form (CRFS9_01B) identified among MSM in China. AIDS Res Hum Retroviruses. 2013.
6 Li Y, Feng Y, Li F, Xue Z, Hu J, Xing H, et al. Genome sequence of a novel HIV-1 circulating recombinant form (CRF79_0107) identified from Shanxi, China. AIDS Res Hum Retroviruses. 2017;33(10):1056–60.
7 Dong Z, Shen Z, Xiong R, Liang F, Liang S, Li J, et al. Near full-length genomic characterization of a novel HIV type 1 CRF01_AE/CRF07_BC recombinant form transmitted between a heterosexual couple in Guangxi, China. AIDS Res Hum Retroviruses. 2016;32(7):689–93.
8 Zhou Y, Li X, Zhang C, Tan G, Stanton B, Zhang X, et al. Rates of HIV, syphilis, and HCV infections among different demographic groups of female sex workers in Guangxi China: evidence from 2010 national sentinel surveillance data. AIDS Care. 2013;25(11):1433–41.
9 Li J, Feng Y, Shen Z, Li Y, Tang Z, Xiong R, et al. HIV-1 transmissions among recently infected individuals in Southwest China are predominantly derived from circulating local strains. Sci Rep. 2018;8(1):12831.
10 Cheng H, Zhang J, Capizzi J, Young NL, Mastro TD. HIV-1 subtype E in Yunnan, China. Lancet. 1994;344(8927):953–4.
11 Liao H, Tee KK, Hase S, Uenishi R, Li XJ, Kusagawa S, et al. Phylogenetic analysis of the dissemination of HIV-1 CRF01_AE in Vietnam. Virology. 2009;391(1):51–6.
12 Su L, Graf M, Zhang Y, von Briesen H, Xing H, Köstler J, et al. Characterization of a virtually full-length human immunodeficiency virus type 1 genome of a prevalent intersubtype (C/B) recombinant strain in China. J Virol. 2000;74(23):11367–76.
13 Meng Z, Xin R, Zhong P, Zhang C, Abubakar YF, Li J, et al. A new migration map of HIV-1 CRF07_BC in China: analysis of sequences from 12 provinces over a decade. PLoS One. 2012;7(12):e52373.
14 Li L, Chen L, Liang S, Liu W, Li T, Liu Y, et al. Subtype CRF01_AE dominate the sexually transmitted human immunodeficiency virus type 1 epidemic in Guangxi, China. J Med Virol. 2013;85(3):388–95.
15 Jiang J, Liang B, Li K, Yang Y, Yang Y, Ning C, et al. Genomic characterization of a novel HIV type 1 strain originating from CRF07_BC and CRF01_AE by heterosexual transmission in the Lingshong prefecture of Guangxi Province, China. AIDS Res Hum Retroviruses. 2020;36(2):153–60.
16 Kong D, Wang Y, Wang C, Liang S, Feng Y, Ruan Y, et al. Characterization of a new HIV-1 CRF01_AE/CRF07_BC recombinant virus in Guangxi, China. AIDS Res Hum Retroviruses. 2017;33(11):1166–70.
17 Liu W, Feng Y, Wang C, Wang Y, Kong D, Qu S, et al. Identification of a novel HIV type 1 CRF01_AE/CRF07_BC recombinant virus in men who have sex with men in GuangXi, China. AIDS Res Hum Retroviruses. 2019;35(4):402–6.
18 Feng Y, He X, Hsi JH, Li F, Li X, Wang Q, et al. The rapidly expanding CRF01_AE epidemic in China is driven by multiple lineages of HIV-1 viruses introduced in the 1990s. *AIDS*. 2013;27(11):1793–802.

19 Cao ZQ, Yang WM, Zhu QY, Lan GH, Shen ZY, Liang SS, et al. [HIV genetic subtypes and comparison of the first CD4(+) T cell counts in newly diagnosed HIV infected patients in Liuzhou, 1998-2012]. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2019;40(5):580–4.

20 Beyrer C, Razak MH, Lisam K, Chen J, Lui W, Yu XF. Overland heroin trafficking routes and HIV-1 spread in south and south-east Asia. *AIDS*. 2000;14(1):75–83.

21 Des Jarlais DC, Friedman SR, Stoneburner RL. HIV infection and intravenous drug use: critical issues in transmission dynamics, infection outcomes, and prevention. *Rev Infect Dis*. 1988;10(1):151–8.