Preplanned Studies

# Characterization of a New HIV-1 Circulating Recombinant Form CRF142_BC — Yunnan, China, 2015 and 2021

Min Chen[1,&]; Huichao Chen[2,&]; Shouxiong Lei[3,&]; Jie Dai[4]; Yanling Ma[2]; Manhong Jia[4,#]


## Summary

### What is already known about this topic?

Recombinant strains dominate the human immunodeficiency virus 1 (HIV-1) epidemic in China. Yunnan Province was the first region in China to report HIV-1 infections in batches. The long-term HIV-1 epidemic led to the generation of various recombinant forms. Among the 47 circulating recombinant forms (CRFs) reported in China, more than 20 were first identified in Yunnan Province.

### What is added by this report?

This study reported a previously unrecognized HIV-1 CRF (CRF142_BC) characterized by the insertion of four short subtype B fragments into the subtype C backbone. CRF142_BC was estimated to have emerged in the mid-1990s, close to the time of the emergence of most known CRF_BCs in China.

### What are the implications for public health practice?

The discovery of new CRFs will provide a basis for HIV-1 molecular tracing and intervention research. In addition, HIV-1 recombination can alter viral biological properties. The study of HIV-1 gene variants needs to be intensified.


Human immunodeficiency virus 1 (HIV-1) is known for its frequent mutation and recombination, which contribute to the genetic diversity of the virus (*1*). Recombination between subtypes of HIV-1 generates unique recombinant forms (URFs) in individuals, some of which become circulating recombinant forms (CRFs) once they circulate in the population. Yunnan Province was the first location in China where the main HIV-1 strains were found. Due to the long-term HIV-1 epidemic, various recombinant forms were generated, including the predominant HIV-1 CRFs currently circulating in China (*2*). In a previous study, this study's authors found five samples that differed from known subtypes/CRFs. To determine whether they were potential CRFs and how they arose, the authors amplified and sequenced the near-full-length genome (NFLG) sequences and performed phylogenetic, recombination, and evolutionary analyses. Phylogenetic analysis revealed that these five sequences clustered in a distinct monophyletic clade distantly related to all known HIV-1 CRFs. Recombination analysis revealed that they shared a subtype C backbone with four subtype B insertions, forming nine subregions. Phylogenetic analyses of the subregions confirmed the parental lineage of each subregion. According to the naming criteria, the strains were named CRF142_BC. Bayesian evolutionary analysis revealed that the time of origin of CRF142_BC was approximately 1994–1997. The identification of new CRFs will provide a basis for the molecular tracing of HIV-1 in China, as well as for the study of HIV mutations and vaccines.

Four of the five samples (15R176, 21ZT314, 21ZT323, and 21ZT334) were collected from Zhaotong City, and one (15R297) was collected from Pu'er City. Two samples (15R176 and 15R297) were collected in 2015, and the other three (21ZT314, 21ZT323, and 21ZT334) were collected in 2021. Demographic information is presented in Table 1. No epidemiological link was found between the five individuals. Written informed consent was obtained from the participants. NFLG sequences were amplified and sequenced as previously described (*3*). Amplified products were sent to SinoGenoMax Co. (Beijing, China) for Sanger sequencing. Phylogenetic analyses were performed using MEGA 11. Recombination breakpoints were analyzed using SimPlot 3.5.1 software. The time of the most recent common ancestor (tMRCA) of CRF142_BC was estimated by Bayesian Markov chain Monte Carlo (MCMC) analysis using BEAST v 1.8.2. The uncorrelated lognormal relaxed molecular clock was used in combination with the Bayesian skyline coalescent tree priors under the GTR+I+G4 nucleotide substitution model.

The NFLG sequences of strains 15R176, 15R297, 21ZT314, 21ZT323, and 21ZT334 were 8,787

TABLE 1. Demographic characteristics of the HIV-1-infected participants.

| Sequence name | Sampling year | Sex | Age | Ethnic group | Education | Marital status | Infection route |
|---|---|---|---|---|---|---|---|
| 15R176 | 2015 | Female | 59 | Han | Illiterate | Married | Heterosexual contact |
| 15R297 | 2015 | Male | 32 | Hani | Primary school | Devoiced/Widowed | Heterosexual contact |
| 21ZT314 | 2021 | Male | 75 | Han | Illiterate | Married | Heterosexual contact |
| 21ZT323 | 2021 | Female | 64 | Miao | Illiterate | Devoiced/Widowed | Heterosexual contact |
| 21ZT334 | 2021 | Male | 36 | Han | Primary school | Devoiced/Widowed | Heterosexual contact |

Abbreviation: HIV-1=human immunodeficiency virus 1.

(640–9,465 in HXB2), 8,795 (670–9,480 in HXB2), 8,814 (635–9,513 in HXB2), 8,924 (637–9,571 in HXB2), and 8,958 (640–9,558 in HXB2) nt in size, respectively, ranging from the 5' noncoding region (NCR) to part of the 3' long terminal repeat (LTR). They were submitted to GenBank under accession numbers PP074169–PP074173.

Phylogenetic analysis revealed that these five sequences formed a distinct monophyletic clade with a bootstrap value of 100%, distantly related to all other HIV-1 subtypes/CRFs (Figure 1A), suggesting they may represent a potential novel CRF. Recombination analysis revealed that these five sequences were composed of subtypes B and C and had similar recombination patterns (Figure 1B). Four segments of subtype B were inserted into the backbone of subtype C, resulting in a mosaic structure of nine subregions: $I_C$ (640-1,555), $II_B$ (1,556-1,690), $III_C$ (1,691-2,885), $IV_B$ (2,886-3,141), $V_C$ (3,142-6,035), $VI_B$ (6,036-6,178), $VII_C$ (6,179-8,865), $VIII_B$ (8,866-9,047) and $IX_C$ (9,048-9,513). Phylogenetic analysis of the nine subregions revealed that subregions I, III, V, VII, and IX clustered with their subtype C counterparts, and subregions II, IV, VI, and VIII clustered with their subtype B counterparts (Figure 2A). These new recombinants are, therefore, designated CRF142_BC. As shown in Figure 1B, compared to CRF08_BC, CRF142_BC had a shorter IIB subregion and one more subtype B insertion fragment ($VI_B$ subregion).

To explore the evolutionary history of CRF142_BC, Bayesian evolutionary analysis was performed with combined subtype C regions (I+III+V+VII+IX) and combined subtype B regions (II+IV+VI+VIII). As shown in Figure 2B, the median tMRCAs of the combined subtype C and subtype B regions were 1994.2 [95% highest probability density (*HPD*): 1989.3–1998.4] and 1997.7 (95% *HPD*: 1976.4–2007.6), respectively, suggesting that CRF142_BC originated between approximately 1994 and 1997. The analysis also showed that the subtype B and subtype C segments were most likely from the Thai B and India C lineages, respectively.

## DISCUSSION

The presence of HIV-1 recombinants has become an important factor in the global epidemic. According to a retrospective analysis, recombinant strains caused 22.8% of HIV-1 infections from 2010 to 2015 (*4*). In East and Southeast Asia, recombinants contribute to more than 80% of infections (*5*). Yunnan Province, bordering Myanmar, Laos, and Vietnam, has historically been on the drug transport route from the "Golden Triangle" drug bases to China's interior, making it the first province in China where HIV-infected individuals were found in groups (*2*). In this epidemic context, subtype B from Thailand and subtype C from India were first introduced in Yunnan through IDUs in the late 1980s (*2*). In the 1990s, these subtypes underwent high-frequency recombination in IDUs, resulting in various BC recombinants (*6*–*7*). Some became CRFs, such as CRF07_BC and CRF08_BC, the world's first CRF_BCs. Afterward, researchers identified a series of CRF_BCs from Yunnan, including CRF57_BC, CRF62_BC, CRF64_BC, CRF86_BC, CRF88_BC, CRF110_BC, and CRF118_BC. Evolutionary studies suggested that these CRF_BCs mostly originated in the 1990s (*6*). Additionally, the subtype C and subtype B segments in CRF_BC were related to India C and Thailand B, respectively (*8*). With the shift in the primary transmission route, CRF_BCs have moved from IDUs to the heterosexual sexually transmitted population. This study found that CRF142_BC also emerged in the 1990s, and its structural fragments were similar in origin to previous CRF_BCs. Additionally, a previous study showed a large number of URFs with similar origins (*9*). These findings suggest that a viral reservoir of BC recombination formed under specific conditions in the 1990s. Some of these recombinant BC strains may have become extinct, while others may still be circulating in the

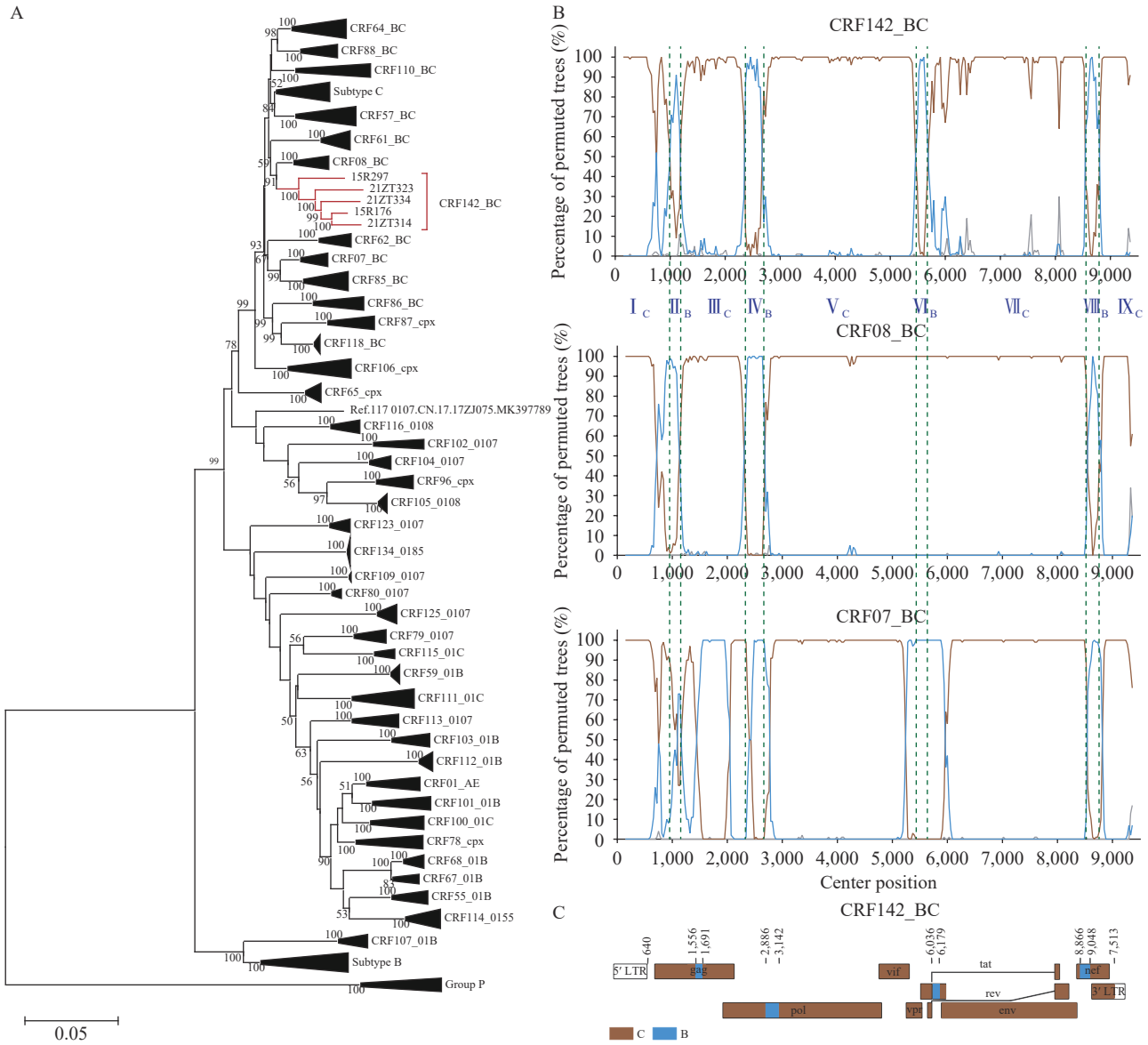


FIGURE 1. Phylogenetic and recombinant analyses based on the near-full-length genome sequence of CRF142_BC. (A) The neighbor-joining phylogenetic tree of the representative HIV-1 CRFs reference sequences. (B) Bootscanning analysis of CRF142_BC, CRF08_BC, and CRF07_BC. (C) Genomic structure of CRF142_BC.
Note: For (A), the sequences of the potential novel CRFs (15R176, 15R297, 21ZT314, 21ZT323, and 21ZT334) are marked in red. The values on the branches represent the percentages of 1,000 bootstrap replicates. The scale bar indicates 5% nucleotide sequence divergence. For (B), conditions used for this analysis were as follows: window: 300 bp, step: 30 bp, GapStrip: on, replicates: 100, Kimura (2-parameter), T/t: 2.0. The Subtype C reference group included AF067155, AF067158, and AF067157. The Subtype B reference group included AY173951, JF932495, and JF932496. The CRF08_BC reference group included KC914396, HM067748, and AY008715. The CRF07_BC reference group included EF368372, EF368370, and AF286230. The reference group of Subtype A4 included AM000053 and AM000054. For (C), the mosaic map was generated using the Recombinant HIV-1 Drawing Tool.
Abbreviation: CRF=circulating recombinant form; LTR= long terminal repeat.

population. Other studies by this study's authors also suggest that there might be other potential CRF_BCs that have yet to be identified.

Among the CRF_BCs found in Yunnan Province, CRF07_BC and CRF08_BC became the most prevalent strains. This dominance may be partly attributable to their earlier emergence, affording them more time to disseminate throughout the population. Second, the biological characteristics of these strains, such as the enhanced transmissibility and decreased virulence reported for CRF07_BC (*10*), may confer a transmission advantage. Third, specific transmission

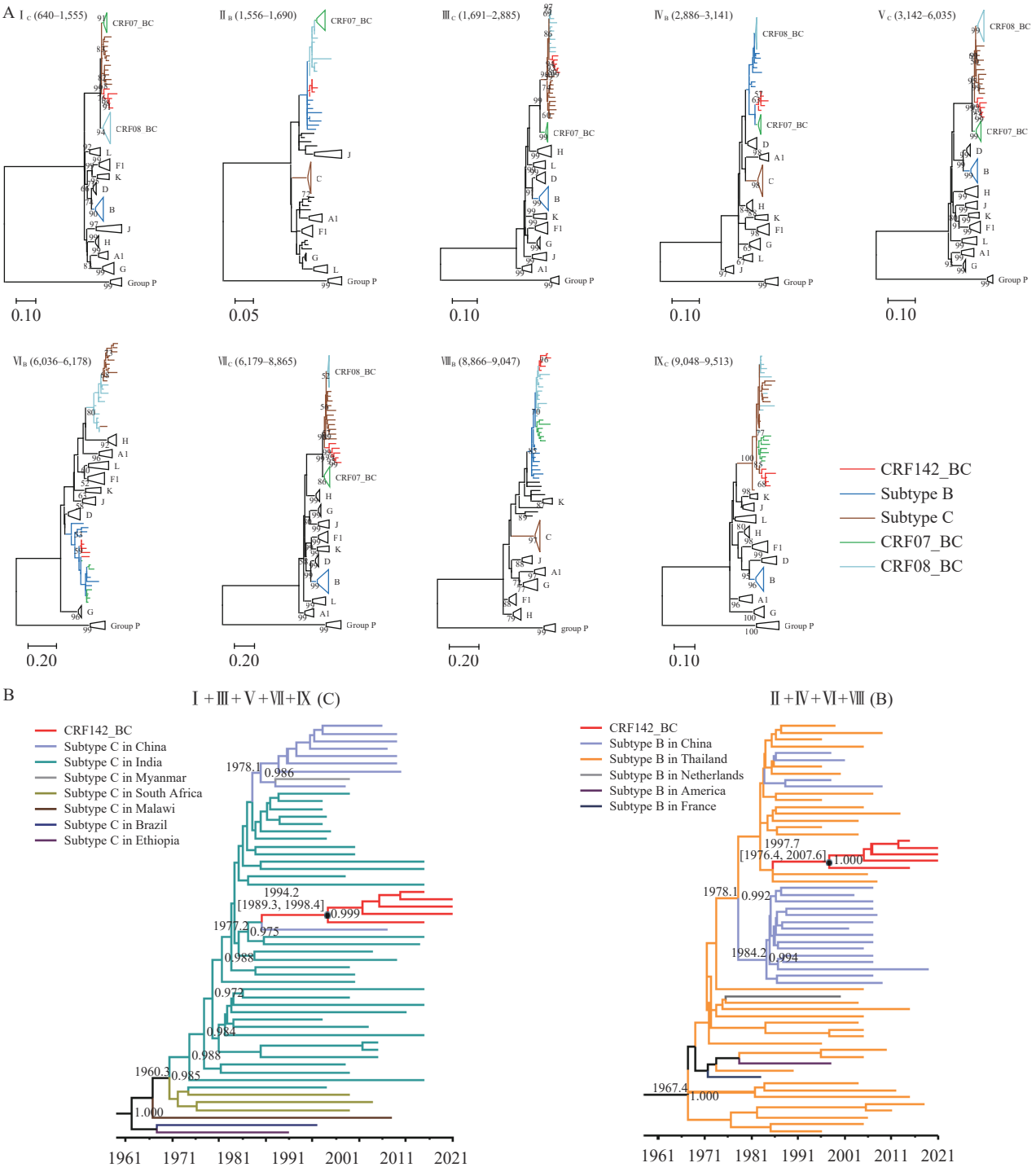


FIGURE 2. Phylogenetic and evolutionary analysis of subregions from CRF142_BC. (A) Maximum likelihood trees of the nine mosaic fragments identified by recombination analysis. (B) The maximum clade credibility (MCC) trees of the combined subtype C subregions (I+III+V+VII+IX) and the combined subtype B subregions (II+IV+VI+VIII) from CRF142_BC.
Note: For (A), the reliability of the tree branches was assessed by 1,000 bootstrap replicates. The scale bar indicates the nucleotide sequence divergence.
Abbreviation: CRF=circulating recombinant form.

events may have contributed to their spread. For example, this study's authors' previous phylodynamic study identified two exponential growth phases for the effective population size of CRF08_BC from 1994 to 1998 and 2001 to 2002 (*11*). The second growth phase appears to have propelled CRF08_BC to

dominance in Yunnan.

In this study, the recombination structures of CRF142_BC and CRF08_BC, which cluster closely in the evolutionary tree, were further compared. Similarities were observed in their recombination patterns, with differences mainly in the $II_B$ and $VI_B$ subregions. The $II_B$ subregion of CRF142_BC was shorter than that of CRF08_BC, while CRF142_BC contained an additional subtype B insertion fragment ($VI_B$ subregion). Therefore, differentiating these CRFs using only partial gene fragments, such as the *pol* gene where CRF142_BC and CRF08_BC share the same recombinant pattern, is difficult. This may explain why CRF142_BC has not been distinguished in previous studies.

The identification of CRF_BCs provides valuable information for HIV-1 genotyping and transmission traceability. In previous studies, some branches on the periphery of the CRF08_BC clade were ambiguously genotyped. However, these poorly genotyped strains were later confirmed to be CRFs, such as CRF57_BC, CRF61_BC, CRF64_BC, CRF88_BC, and CRF110_BC. Characterization of BC recombination will further improve the resolution of HIV-1 genotyping, allowing for more precise determination of HIV-1 subtypes and CRFs. This information can be used to track the transmission of HIV-1, enabling more targeted prevention and control efforts. More importantly, recombination that occurs in specific regions, such as genes encoding outer membrane proteins, metabolic enzymes, and expression regulators, may alter the biological properties of the virus, such as immunogenicity, drug resistance, replication, and fitness (*12*–*13*). Therefore, continuous monitoring and research on HIV-1 recombinants are essential to improve our understanding of their biological properties and transmission patterns and to develop effective strategies for controlling their spread.

This study also has some limitations. According to the nomenclature rule, only three epidemiologically unrelated NFLG sequences are required to identify a new CRF. Although five such sequences were reported in this study, this limited number is insufficient to fully understand the variation and evolutionary characteristics of the strain, and further surveillance is needed to fully understand its epidemiological characteristics and trends in the population. Secondly, this study focuses mainly on the genomic characteristics of the virus, and its biological properties have not been thoroughly investigated.

In conclusion, this study reports a novel HIV-1 CRF, designated CRF142_BC, identified in Yunnan Province. This CRF is characterized by inserting four subtype B fragments into a subtype C backbone. This study estimates that CRF142_BC originated in the mid-1990s, a timeframe similar to the tMRCAs of most CRF_BCs identified to date. The identification of this CRF has implications for understanding HIV-1 variation and evolution and will facilitate more refined HIV-1 genotyping in clinical practice.

**Conflicts of interest**: No conflicts of interest.

**Funding:** Supported by the National Natural Science Foundation of China (82160635) and the Yunnan Revitalization Talents Support Program (Special Project for Famous Doctors).

**doi:** 10.46234/ccdcw2024.222

[#] Corresponding author: Manhong Jia, jiamanhong@hotmail.com.

[1] Yunnan Provincial Key Laboratory of Public Health and Biosafety & Health Laboratory Center, Yunnan Center for Disease Control and Prevention, Kunming City, Yunnan Province, China; [2] Institute for AIDS/STD Control and Prevention, Yunnan Center for Disease Control and Prevention, Kunming City, Yunnan Province, China; [3] Division for AIDS/STD Control and Prevention, Zhaotong Center for Disease Control and Prevention, Zhaotong City, Yunnan Province, China; [4] Yunnan Provincial Key Laboratory of Public Health and Biosafety & Institute for AIDS/STD Control and Prevention, Yunnan Center for Disease Control and Prevention, Kunming City, Yunnan Province, China.

[&] Joint first authors.

Submitted: February 20, 2024; Accepted: June 19, 2024

## REFERENCES

1. Smyth RP, Davenport MP, Mak J. The origin of genetic diversity in HIV-1. Virus Res 2012;169(2):415 – 29. https://doi.org/10.1016/j.virusres.2012.06.015.
2. Lu L, Jia MH, Ma YL, Yang L, Chen ZW, Ho DD, et al. The changing face of HIV in China. Nature 2008;455(7213):609 – 11. https://doi.org/10.1038/455609a.
3. Grossmann S, Nowak P, Neogi U. Subtype-independent near full-length HIV-1 genome sequencing and assembly to be used in large molecular epidemiological studies and clinical management. J Int AIDS Soc 2015;18(1):20035. https://doi.org/10.7448/IAS.18.1.20035.
4. Hemelaar J, Elangovan R, Yun J, Dickson-Tetteh L, Kirtley S, Gouws-Williams E, et al. Global and regional epidemiology of HIV-1 recombinants in 1990-2015: a systematic review and global survey. Lancet HIV 2020;7(11):e772 – 81. https://doi.org/10.1016/S2352-3018(20)30252-6.
5. Hemelaar J, Elangovan R, Yun J, Dickson-Tetteh L, Fleminger I, Kirtley S, et al. Global and regional molecular epidemiology of HIV-1, 1990-2015: a systematic review, global survey, and trend analysis. Lancet Infect Dis 2019;19(2):143 – 55. https://doi.org/10.1016/S1473-3099(18)30647-9.
6. Feng Y, Takebe Y, Wei HM, He X, Hsi JH, Li ZP, et al. Geographic origin and evolutionary history of China's two predominant HIV-1 circulating recombinant forms, CRF07_BC and CRF08_BC. Sci Rep 2016;6:19279. https://doi.org/10.1038/srep19279.
7. McCutchan FE, Carr JK, Murphy D, Piyasirisilp S, Gao F, Hahn B, et al. Precise mapping of recombination breakpoints suggests a common parent of two BC recombinant HIV type 1 strains circulating in China.

AIDS Res Hum Retroviruses 2002;18(15):1135 – 40. https://doi.org/10.1089/088922202320567879.
8. Wei HM, Liu YJ, Feng Y, Hsi J, Xing H, He X, et al. Genome sequence of a novel HIV-1 circulating recombinant form (CRF57_BC) identified from Yunnan, China. AIDS Res Hum Retroviruses 2014;30(4):384 – 8. https://doi.org/10.1089/aid.2013.0228.
9. Han XX, An MH, Zhao B, Duan S, Yang SM, Xu JJ, et al. High prevalence of HIV-1 intersubtype B'/C recombinants among injecting drug users in Dehong, China. PLoS One 2013;8(5):e65337. https://doi.org/10.1371/journal.pone.0065337.
10. Cheng ZT, Yan HC, Li QM, Ablan SD, Kleinpeter A, Freed EO, et al. Enhanced transmissibility and decreased virulence of HIV-1 CRF07_BC may explain its rapid expansion in China. Microbiol Spectr 2022;10(4):e0014622. https://doi.org/10.1128/spectrum.00146-22.
11. Chen M, Jia MH, Ma YL, Luo HB, Chen HC, Yang CJ, et al. The changing HIV-1 genetic characteristics and transmitted drug resistance among recently infected population in Yunnan, China. Epidemiol Infect 2018;146(6):775 – 81. https://doi.org/10.1017/S0950268818000389.
12. Ritchie AJ, Cai FP, Smith NM, Chen SR, Song HS, Brackenridge S, et al. Recombination-mediated escape from primary CD8+ T cells in acute HIV-1 infection. Retrovirology 2014;11:69. https://doi.org/10.1186/s12977-014-0069-9.
13. Nora T, Charpentier C, Tenaillon O, Hoede C, Clavel F, Hance AJ. Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. J Virol 2007;81(14):7620 – 8. https://doi.org/10.1128/JVI.00083-07.