

Methods and Applications

Identification of Multiple Unique HIV-1 Recombinant Forms in Newly Reported HIV-1 Infected Individuals — Anhui Province, China, 2023

Huan Li¹; Yi Feng¹; Qi Li¹; Jianjun Wu²; Hui Xing¹; Lingjie Liao¹; Zheng Wang^{1,†}

ABSTRACT

Introduction: The genetic diversity of human immunodeficiency virus-1 (HIV-1) in China is characterized by multiple subtypes, circulating recombinant forms (CRFs) and unique recombinant forms (URFs) across the country. Through timely molecular surveillance, over 65 distinct CRFs have been identified in China to date. In this study, we identified five novel URFs among newly reported HIV-1 infected individuals in Anhui Province, China.

Methods: Near-full length HIV genome sequences were obtained using two-half molecule amplification methods from five samples containing potential URFs. The sequences were subsequently subjected to phylogenetic and recombination analyses.

Results: Phylogenetic and recombination analyses of the five near-full length genome sequences confirmed their classification as novel URFs. Among these, three sequences were recombinants of CRF01_AE and CRF07_BC, one sequence was a recombinant of CRF01_AE, CRF07_BC and B, and one sequence resulted from CRF07_BC and CRF08_BC recombination.

Conclusions: The identification of URFs in newly infected individuals indicates ongoing transmission of multiple HIV-1 clades in Anhui Province, with superinfection occurring at notable frequencies. These findings emphasize the importance of enhancing long-term surveillance of circulating HIV-1 clades using near-full length sequence analysis in Anhui, China.

The first acquired immune deficiency syndrome (AIDS) case in China was reported in 1985 (1), marking the beginning of HIV-1 spread throughout the country. According to the National Statutory Infectious Disease Epidemic Profile published by the National Bureau of Disease Control and Prevention

(NBDCP) in 2023, China documented 59,533 AIDS cases and 22,393 related deaths (<https://www.ndcpa.gov.cn/>). Although the implementation of immediate identification and treatment strategies has significantly reduced HIV incidence, HIV-1 remains a substantial threat to public health in China.

In Anhui Province, the identification of the first human immunodeficiency virus-positive (HIV-positive) case in 1994 was followed by a rapid increase in infections, particularly among commercial blood donors (2). Anhui has served as a crucial hub in the dissemination of HIV-1 across China (3). By 2018, the province had documented 17,183 HIV-infected individuals. The predominant HIV-1 subtypes circulating in Anhui include CRF01_AE, CRF07_BC, CRF08_BC, CRF55_01B, and subtype B. Previous studies have revealed a significant prevalence of unique recombinant forms (URFs), indicating frequent superinfections. The ongoing recombination between subtypes enhances HIV genetic diversity and facilitates the emergence of highly adaptive variants that could potentially trigger new pandemics (4). Given the substantial subtype diversity in Anhui, continuous monitoring of subtype distributions remains essential for effective HIV containment (5). Furthermore, surveillance of HIV genetic characteristics through near-full length genome (NFLG) analysis provides crucial insights into the HIV epidemic in Anhui and informs the development of innovative strategies for preventing and treating new HIV infections.

METHODS

Study Population

The sampling strategy in Anhui Province employed a risk-stratified approach, categorizing prefectural-level cities into high, medium, and low risk based on reported HIV cases over the preceding three years. The sampling framework mandated the inclusion of at least two cities from each risk category, with minimum

sampling proportions of 20%, 40%, and 60% for high-, medium-, and low-risk cities, respectively.

The study encompassed 238 newly confirmed HIV-1 positive samples collected from Anhui in 2023. Plasma samples were procured through the Anhui Provincial CDC.

Sequencing and Analysis of PR-RT, IN, and *env* Fragments

Viral RNA was extracted from plasma samples using the QIAamp Viral RNA Mini Kit (QIAGEN, Germany). HIV-1 cDNA synthesis and amplification were performed using the AccessQuick™ RT-PCR System (PROMEGA, China) in combination with 2× Taq PCR Mix (TIANGEN, China). The amplification and sequencing of HIV genome fragments were conducted according to previously established protocols (6–7).

Sequencing was performed on an ABI 3730XL sequencer (Applied Biosystems, USA). Sequence processing, including trimming, splicing, and mixed base interpretation, was conducted using Sequencher v5.4.6 (Gene Codes Corporation, Mississippi, USA). To ensure data quality and exclude experimental cross-contamination, sequences were validated using the WHO HIVDR QC Tool. Multiple sequence alignment was performed using MAFFT v7 (Research Institute for Microbial Diseases Osaka University, Osaka, Japan) (8). Reference sequences were obtained from relevant databases, and phylogenetic analysis was conducted using IQ-TREE (version 2.0, University of Vienna, Vienna, The Republic of Austria; The Australian National University, Canberra, Australia) (9) with the Maximum Likelihood (ML) method, implementing the General Time Reversible (GTR) + G nucleotide substitution model and 1,000 bootstrap replicates. The resulting phylogenetic trees were visualized using the interactive tree of life (iTOL) online platform (<https://itol.embl.de/>).

Sequencing and Analysis of Near Full-length Genome

Near full-length genome sequences were obtained using a two-amplicon strategy following previously established protocols. The methodologies for viral RNA extraction, PCR amplification, and sequencing procedures have been detailed in previous publications (10).

Sequence analysis was performed using Sequencher software (version 5.4.6, Gene Codes Corporation,

USA) for sequence trimming, secondary peak detection, sequence assembly, and mixed base interpretation. Quality control of the assembled sequences was conducted using the online Gene Cutter tool (https://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html). To eliminate potential contamination, all sequences were compared against the comprehensive HIV sequence database using the Basic Local Alignment Search Tool (BLAST) (<http://hiv-web.lanl.gov/content/index>). Sequences meeting quality control criteria were analyzed for potential unique recombinant patterns using two complementary approaches: the Recombinant Identification Program (RIP) with default parameters (<https://www.hiv.lanl.gov/content/sequence/RIP/RIP.html>) and the jumping profile Hidden Markov Model (jpHMM) (<http://jphmm.gobics.de/>) within the HIV Database.

Based on the RIP and jpHMM analyses, we constructed a comprehensive reference sequence alignment incorporating predominant subtypes circulating in China (CRF01_AE, CRF07_BC, CRF08_BC, CRF55_01B, and subtype B), other Chinese CRFs, and relevant external sequences. Sequence alignment was performed using Aliview software (Uppsala University, Uppsala, Sweden) (11). Phylogenetic analysis was conducted using IQ-TREE (version 2.0, University of Vienna, Vienna, The Republic of Austria; The Australian National University, Canberra, Australia) (9) with the GTR model and 1,000 bootstrap replicates. The resulting phylogenetic trees were visualized using the iTOL online platform. Recombination breakpoint analysis was performed using SimPlot (version 3.5.1, Biomatters, Auckland, New Zealand) (12) with parameters set to a 300 bp window size and 20 bp step size. The genetic structure of unique recombinant samples was mapped using the Recombinant Mapping Tool (https://www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html) available in the HIV Database.

RESULTS

Demographic Information of the Study Population

From the total cohort of 238 newly reported cases in Anhui Province, 37 cases (15.5%) demonstrated subtype inconsistencies across at least two of the three analyzed genomic regions: PR-RT (nucleotides 2253-

3500, HXB2), IN (nucleotides 4230-5096, HXB2), and *env* (nucleotides 7016-7650, HXB2). Among these 37 samples exhibiting subtype discordance, seven contained sufficient plasma volume for NFLG amplification and sequencing, ultimately yielding 5 complete NFLG sequences. These 5 sequences were confirmed as URFs and designated as AHAQ230009, AHBZ230009, AHBZ230031, AHCUZ230011, and AHXC230017. The demographic characteristics of these 5 individuals are detailed in [Supplementary Table S1](#) (available at <https://weekly.chinacdc.cn/>).

Phylogenetic Analysis of Genomic Regions

Comprehensive phylogenetic analysis was performed on three distinct genomic regions: PR-RT, IN, and *env*. Reference sequences were obtained from HIV databases, and maximum likelihood (ML) phylogenetic trees were constructed using IQ-TREE (version 2.0), implementing the GTR + G nucleotide substitution model with 1,000 bootstrap replicates. Tree visualization was accomplished using the iTOL online

web tool. The subtype classifications for the three regions (PR-RT, IN, and *env*) corresponding to the five sequences are summarized in [Figure 1](#) and [Table 1](#).

Near Full-length Genome Analysis

The sequences obtained from the five potential URFs exceeded 8,000 base pairs in length, encompassing the complete coding regions for *gag*, *pol*, *vif*, *vpr*, *vpu*, *rev*, *tat*, *env*, and *nef* genes. Phylogenetic analysis conducted using IQ-TREE2 ([Figure 2](#)) demonstrates that these five samples represent distinct lineages independent of previously identified CRFs.

Detailed analysis revealed specific breakpoint locations within the HXB2 genome reference framework for all five sequences. As shown in [Figure 3](#), the mosaic structures of these recombinants are characterized as follows: AHAQ230009 comprises three segments: I_{CRF07_BC} (634-4884); II_{CRF01_AE} (4885-8404); and III_{CRF07_BC} (8405-8960). AHBZ230009 consists of three regions: I_{CRF07_BC} (634-6323); II_{CRF01_AE} (6324-8327); and III_{CRF07_BC} (8328-9552). AHBZ230031 exhibits a

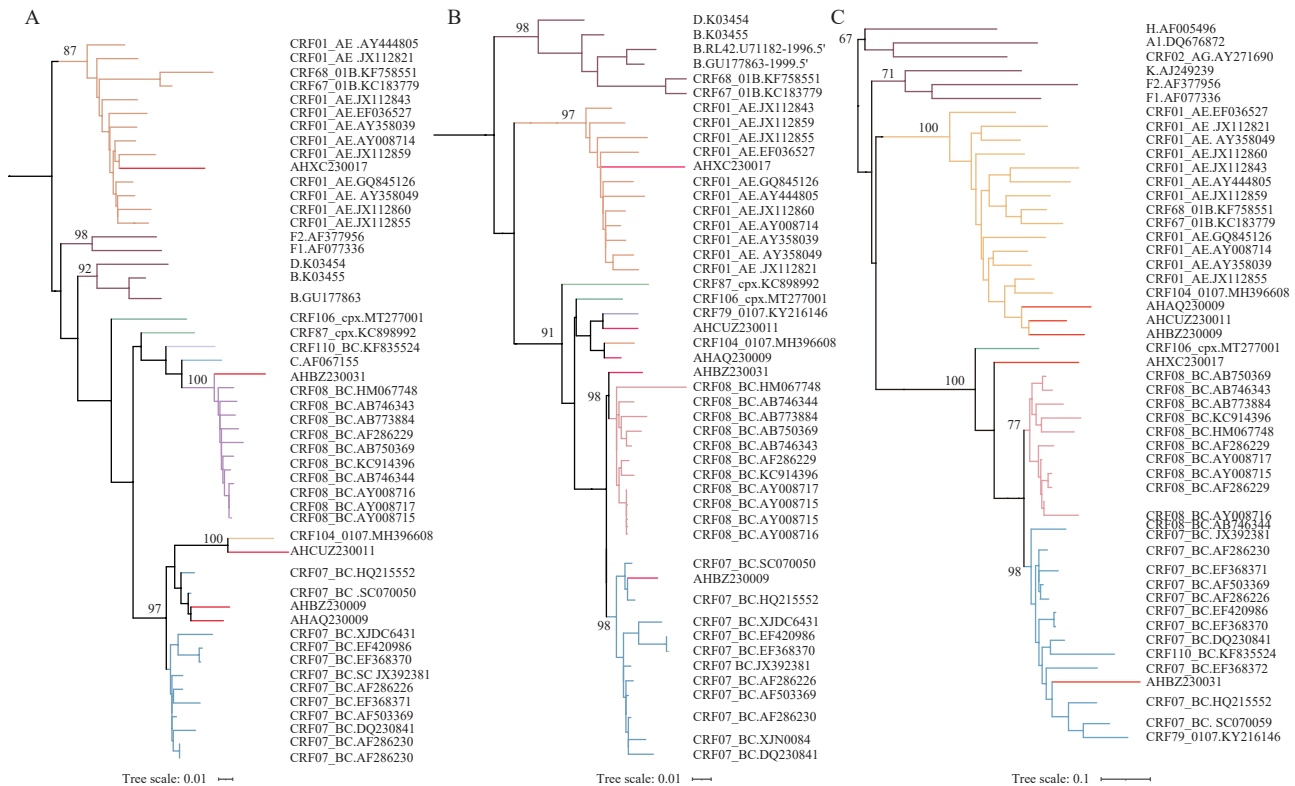


FIGURE 1. Phylogenetic tree of three fragments of five samples. (A) PR-RT; (B) IN; (C) *env*.

Note: Phylogenetic tree produced by IQ-TREE, the reliability of the tree branches was assessed by 1,000 bootstrap replicates, and trees were visualized by iTOL. The red line represents the five samples (AHAQ230009, AHBZ230009, AHBZ230031, AHCUZ230011 and AHXC230017).

Abbreviation: PR-RT=protease-reverse transcriptase; IN=integrase; *env*=envelope; iTOL=the interactive tree of life.

TABLE 1. Subtypes of three fragments (PR-RT, IN, *env*) of five samples.

Sample ID	Subtype of PR-RT	Subtype of IN	Subtype of <i>env</i>
AHAQ230009	CRF07_BC	CRF104_0107	CRF01_AE
AHBZ230009	CRF07_BC	CRF07_BC	CRF01_AE
AHBZ230031	CRF08_BC	CRF08_BC	CRF07_BC
AHCUZ230011	CRF104_0107	CRF79_0107	CRF01_AE
AHXC230017	CRF01_AE	CRF01_AE	URF

Abbreviation: PR-RT=protease-reverse transcriptase; IN=integrase; *env*=envelope; URF=unique recombinant forms.

more complex structure with seven segments: I_{CRF07_BC} (620-1933); II_{CRF08_BC} (1934-3001); III_{CRF07_BC} (3002-3215); IV_{CRF08_BC} (3216-4628); V_{CRF07_BC} (4629-8518); VI_{CRF08_BC} (8519-9098); and VII_{CRF07_BC} (9099-9470). AHCUZ230011 displays six distinct regions: I_{CRF07_BC} (608-2959); II_{CRF01_AE} (2960-4244); III_{CRF07_BC} (4245-4885); IV_{CRF01_AE} (4886-8327); V_{CRF07_BC} (8328-8811); and VI_{CRF01_AE} (8812-9354). AHXC230017 shows six segments: I_{CRF01_AE} (621-4832); II_B (4833-4899); III_{CRF01_AE} (4900-6937); IV_{CRF07_BC} (6938-7679); V_B (7680-8240); and VI_{CRF01_AE} (8241-9552) (Supplementary Table S2, available at <https://weekly.chinacdc.cn/>). The phylogenetic analyses for all identified fragments are presented in Supplementary Figure S1 (available at <https://weekly.chinacdc.cn/>).

Bootscan analysis (Figure 3A) and similarity plot (Supplementary Figure S2, available at <https://weekly.chinacdc.cn/>) provide substantial evidence that these five isolates represent unique recombinants, distinct from previously identified CRFs or URFs. The recombinant structures of the five near-full-length sequences were delineated based on the breakpoint locations (Figure 3B), thereby confirming their classification as novel URFs.

Drug Resistance Analysis

Analysis of drug resistance mutations revealed that none of the five NFLGs exhibited resistance-associated mutations in either the protease or integrase regions. However, the AHXC230017 sequence contains the V179D mutation in the reverse transcriptase region, which is associated with non-nucleoside reverse transcriptase inhibitors. This mutation conferred intermediate-level resistance to Efavirenz (EFV) and Nevirapine (NVP), while demonstrating low-level resistance to Rilpivirine (RPV).

DISCUSSIONS

In this study, we identified five URFs in Anhui

Province through comprehensive NFLG analysis. These URFs exhibited distinct recombination patterns, characterized by mosaic segments derived from multiple viral subtypes distributed throughout their genomes. While current CRF classification primarily relies on NFLG sequence analysis (13–14), this approach offers significant advantages over shorter gene fragment analysis (such as *PR-RT*, *IN*, and *env*), particularly in delineating recombination patterns and precise breakpoint locations. This methodology provides crucial insights for HIV-1 molecular epidemiological surveillance.

Recent years, the rapid dissemination of HIV-1 CRF07_BC strain among sexually transmitted populations in China leads to the emergence of numerous variants, including both URFs and novel CRFs (15). Our finding that CRF07_BC participated in all five URF recombination events suggests its enhanced transmission capability and infection prevalence (1).

Anhui Province, strategically located in central China and bordered by Jiangsu, Zhejiang, and Shanghai, serves as a major labor-exporting inland province and has emerged as a significant hub for HIV transmission and spread (3). The co-circulation of multiple subtypes in this region increases the probability of inter-subtype recombination (16), potentially challenging existing prevention and treatment strategies. Therefore, vigilant monitoring of novel URFs is essential for effectively evaluating and providing early warning of emergent strains with enhanced transmission potential.

However, the relatively small sample size of this study limits its contribution to monitoring the epidemic situation across the entire province. Furthermore, this study employed Sanger sequencing, which may not identify occurrences of superinfection or drug resistance mutations in certain samples. Future efforts should concentrate on increasing the sample size and utilizing more advanced sequencing techniques such as Nanopore technology to analyze the samples,

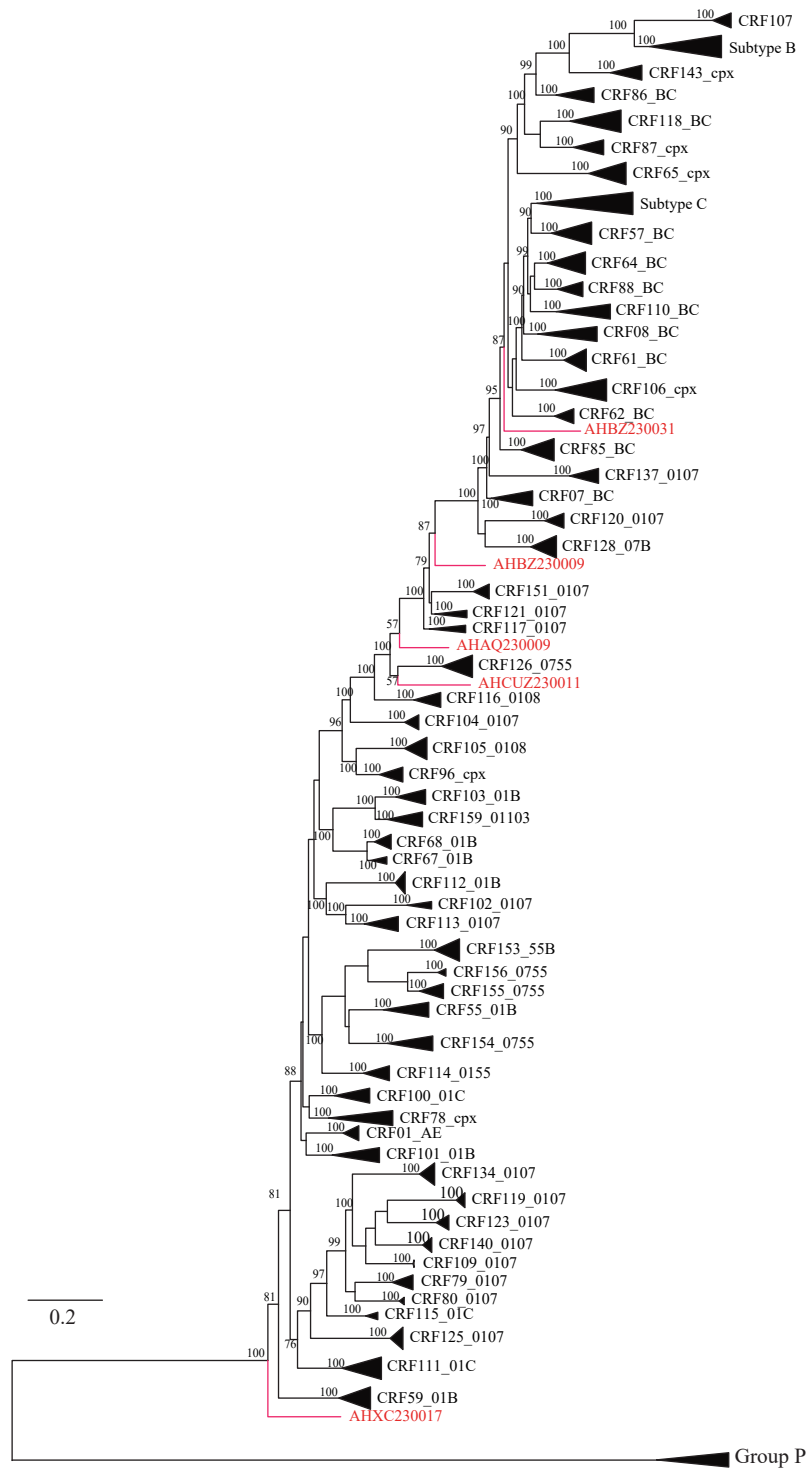


FIGURE 2. Phylogenetic tree of five NFLGs.

Note: The sequences of the potential URFs are marked in red. The neighbor joining tree of five NFLGs was constructed using IQ-TREE 2, and the reliability of the tree branches was assessed by 1,000 bootstrap replicates, and trees were visualized by iTOL.

Abbreviation: NFLGs=near-full length genome; URF=unique recombinant forms; iTOL=the interactive tree of life.

thereby obtaining more comprehensive information.

In conclusion, our identification of five URFs through near full-length HIV-1 genome analysis not

only expands our understanding of HIV-1 genetic diversity but also underscores the critical importance of NFLG sequencing in molecular surveillance efforts.

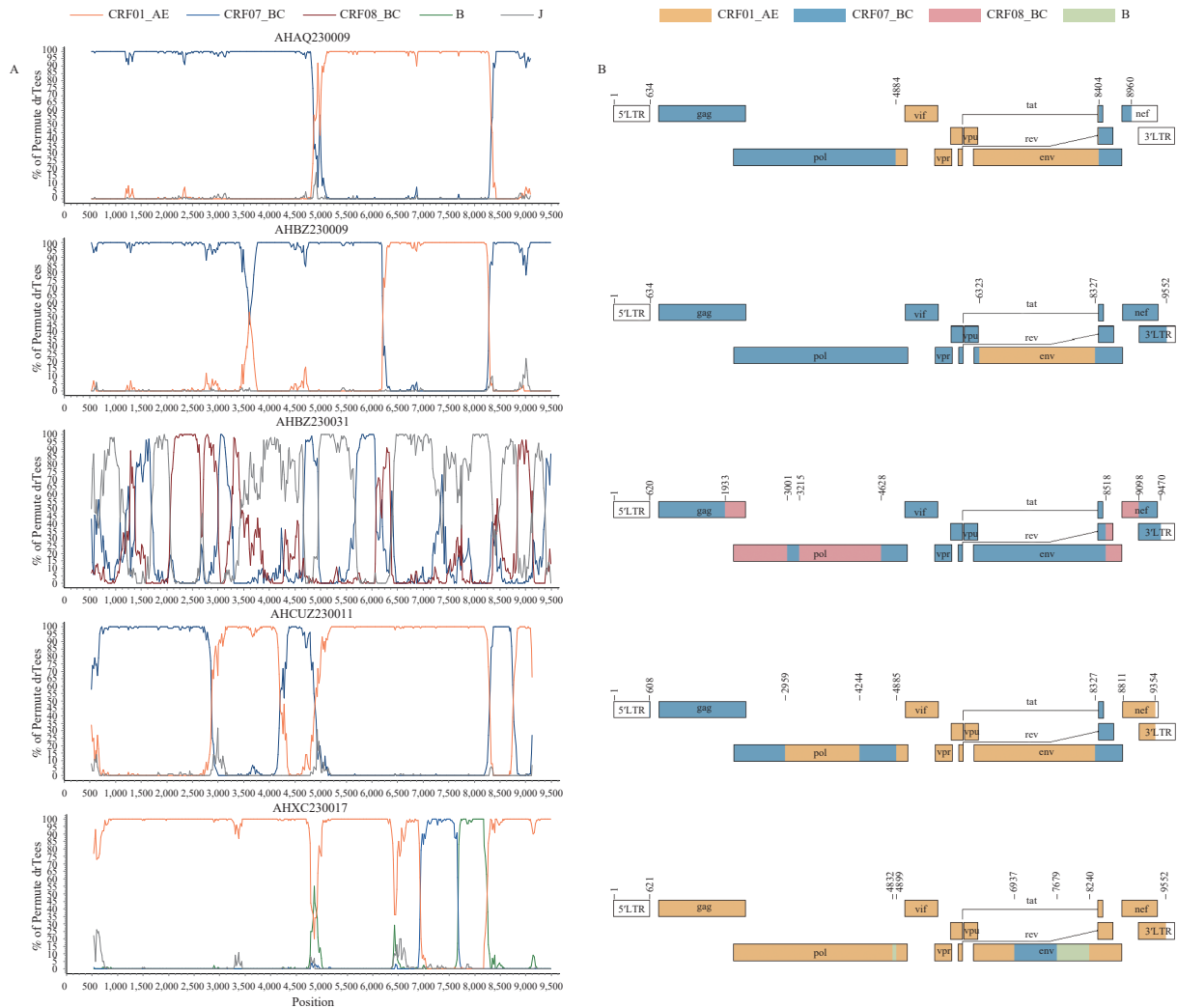


FIGURE 3. Recombinant analysis of the five identified URFs. (A) Bootscan analyses of the five URFs; (B) Genome map of the five URFs.

Note: (A) Bootscan analyses of the five URFs were performed using a window size of 300 bases and a step size of 20 bases, GapStrip: on, Kimura (2-parameter), T/t: 4.0. The CRF01_AE reference group includes JX112859, EF036527, and AY358049. The subtype CRF07_BC reference group includes SC070050, HQ215552, AF286230. The subtype CRF08_BC reference group includes KC914396 and AY008715. The subtype B reference group includes GU177863 and U71182. The subtype J reference is SE9280. The x-axis represents the nucleotide positions, while the y-axis of the bootscan analysis shows the percentage bootstrap values of the permuted trees. (B) The analysis of recombination breakpoints in the five URFs. The mosaic fragments in the near-full-length genomes (NFLGs) recombinants are color-coded as follows: CRF01_AE (orange), CRF07_BC (blue), CRF08_BC (pink), and subtype B (green). The nucleotide positions of each fragment are numbered according to the HIV-1 reference sequence HXB2 (K03455).

Abbreviation: URF=unique recombinant forms.

Conflicts of interest: No conflicts of interest.

Acknowledgements: The staff of the Anhui Provincial CDC for their assistance with sample collection and processing.

Ethical statement: All participants provided informed consent. Approved by the Ethics Committee of the National Center for AIDS/STD Control and Prevention, China CDC (approval number X140617334).

Funding: Supported by the Ministry of Science and Technology of China (grant number 2022YFC2305201).

doi: 10.46234/ccdcw2025.046

Corresponding author: Zheng Wang, wangzheng@chinaaids.cn.

¹ National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China; ² Anhui Provincial Center for Disease Control and Prevention, Hefei City, Anhui Province, China.

Copyright © 2025 by Chinese Center for Disease Control and Prevention. All content is distributed under a Creative Commons Attribution Non Commercial License 4.0 (CC BY-NC).

Submitted: November 02, 2024

Accepted: February 06, 2025

Issued: February 21, 2025

REFERENCES

1. Yang XG, Zhu HY, An WN, Zhao J, Lu XL, Sun WL, et al. Genetic characterization of a novel HIV-1 CRF01_AE/CRF07_BC recombinant form found among men who have sex with men in Baoding City, Hebei Province, China. *Arch Virol* 2022;167(11):2395 – 402. <https://doi.org/10.1007/s00705-022-05563-y>.
2. Zheng S, Wu JJ, Hu ZW, Gan MZ, Liu L, Song C, et al. Epidemiology and molecular transmission characteristics of HIV in the capital city of Anhui Province in China. *Pathogens* 2021;10(12):1554. <https://doi.org/10.3390/pathogens10121554>.
3. Wu JJ, Meng ZF, Xu JQ, Lei YH, Jin L, Zhong P, et al. New emerging recombinant HIV-1 strains and close transmission linkage of HIV-1 strains in the Chinese MSM population indicate a new epidemic risk. *PLoS One* 2013;8(1):e54322. <https://doi.org/10.1371/journal.pone.0054322>.
4. Zhang D, Wu JJ, Zhang Y, Shen YL, Dai SY, Wang XL, et al. Genetic characterization of HIV-1 epidemic in Anhui Province, China. *Virol J* 2020;17(1):17. <https://doi.org/10.1186/s12985-020-1281-y>.
5. Wu JJ, Zhang Y, Shen YL, Wang XL, Xing H, Yang XH, et al. Phylogenetic analysis highlights the role of older people in the transmission of HIV-1 in Fuyang, Anhui Province, China. *BMC Infect Dis* 2019;19(1):562. <https://doi.org/10.1186/s12879-019-4187-9>.
6. Zhou ZH, Ma P, Feng Y, Ou WD, Wei M, Shao YM. The inference of HIV-1 transmission direction between a man who has sex with men and his heterosexual wife based on the sequences of HIV-1 quasi-species. *Emerging Microbes Infect* 2021;10(1):1209 – 16. <https://doi.org/10.1080/22221751.2021.1938693>.
7. Fan WG, Wang XD, Zhang YC, Meng J, Su MM, Yang XG, et al. Prevalence of resistance mutations associated with integrase inhibitors in therapy-naïve HIV-positive patients in Baoding, Hebei Province, China. *Front Genet* 2022;13:975397. <https://doi.org/10.3389/fgene.2022.975397>.
8. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 2019;20(4):1160 – 6. <https://doi.org/10.1093/bib/bbx108>.
9. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;37(5):1530 – 4. <https://doi.org/10.1093/molbev/msaa015>.
10. Marichannegowda MH, Setua S, Bose M, Sanders-Buell E, King D, Zemil M, et al. Transmission of highly virulent CXCR4 tropic HIV-1 through the mucosal route in an individual with a wild-type CCR5 genotype. *eBioMedicine* 2024;109:105410. <https://doi.org/10.1016/j.ebiom.2024.105410>.
11. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 2014;30(22):3276 – 8. <https://doi.org/10.1093/bioinformatics/btu531>.
12. Samson S, Lord E, Makarenkov V. SimPlot++: a Python application for representing sequence similarity and detecting recombination. *Bioinformatics* 2022;38(11):3118 – 20. <https://doi.org/10.1093/bioinformatics/btac287>.
13. Chen M, Ma YL, Chen HC, Dai J, Dong LJ, Jia MH. Identification of a newly emerging second-generation HIV-1 circulating recombinant form (CRF145_0755) among men who have sex with men in China. *J Infect* 2024;88(3):106126. <https://doi.org/10.1016/j.jinf.2024.106126>.
14. Chen M, Chen HC, Lei SX, Dai J, Ma YL, Jia MH. Characterization of a new HIV-1 circulating recombinant form CRF142_BC — Yunnan, China, 2015 and 2022. *China CDC Wkly* 2024;6(42):1080 – 5. <https://doi.org/10.46234/ccdcw2024.222>.
15. Liu X, Wang D, Hu J, Song C, Liao LJ, Feng Y, et al. Changes in HIV-1 subtypes/sub-subtypes, and transmitted drug resistance among ART-naïve HIV-infected individuals - China, 2004-2022. *China CDC Wkly* 2023;5(30):664 – 71. <https://doi.org/10.46234/ccdcw2023.129>.
16. Luan H, Han XX, Yu XO, An MH, Zhang H, Zhao B, et al. Dual infection contributes to rapid disease progression in men who have sex with men in China. *J Acquir Immune Defic Syndr* 2017;75(4):480 – 7. <https://doi.org/10.1097/QAI.0000000000001420>.

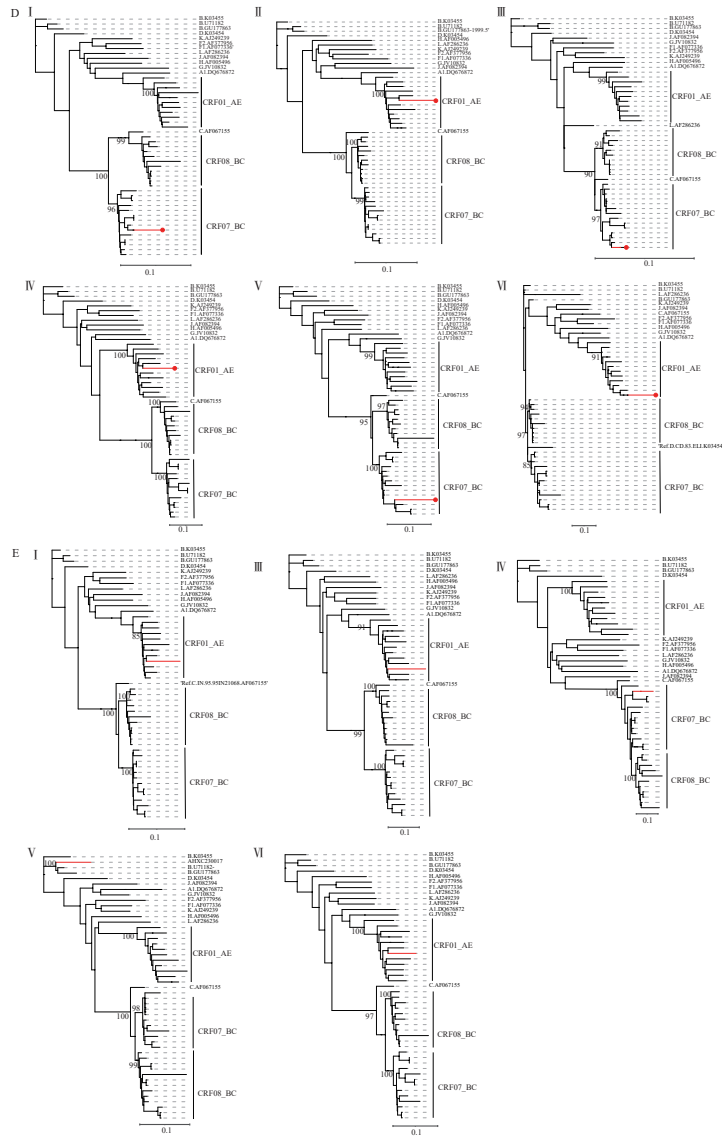
SUPPLEMENTARY MATERIALS

SUPPLEMENTARY TABLE S1. Demographic and clinical characteristics of newly reported cases.

Sample ID	Transmission	Age	Gender	Marriage	CD4+ (cells/mm ³)	Diagnose time	Sampling time
AHAQ230009	HET	31	Male	Unmarried	302	2023.5.8	2023.5.10
AHBZ230009	MSM	35	Male	Married	579	2023.4.20	2023.4.25
AHBZ230031	HET	29	Male	Unmarried	540	2023.5.25	2023.5.30
AHCUZ230011	HET	41	Male	Unmarried	227	2023.5.26	2023.6.13
AHXC230017	HET	56	Female	Married	228	2023.6.19	2023.6.27

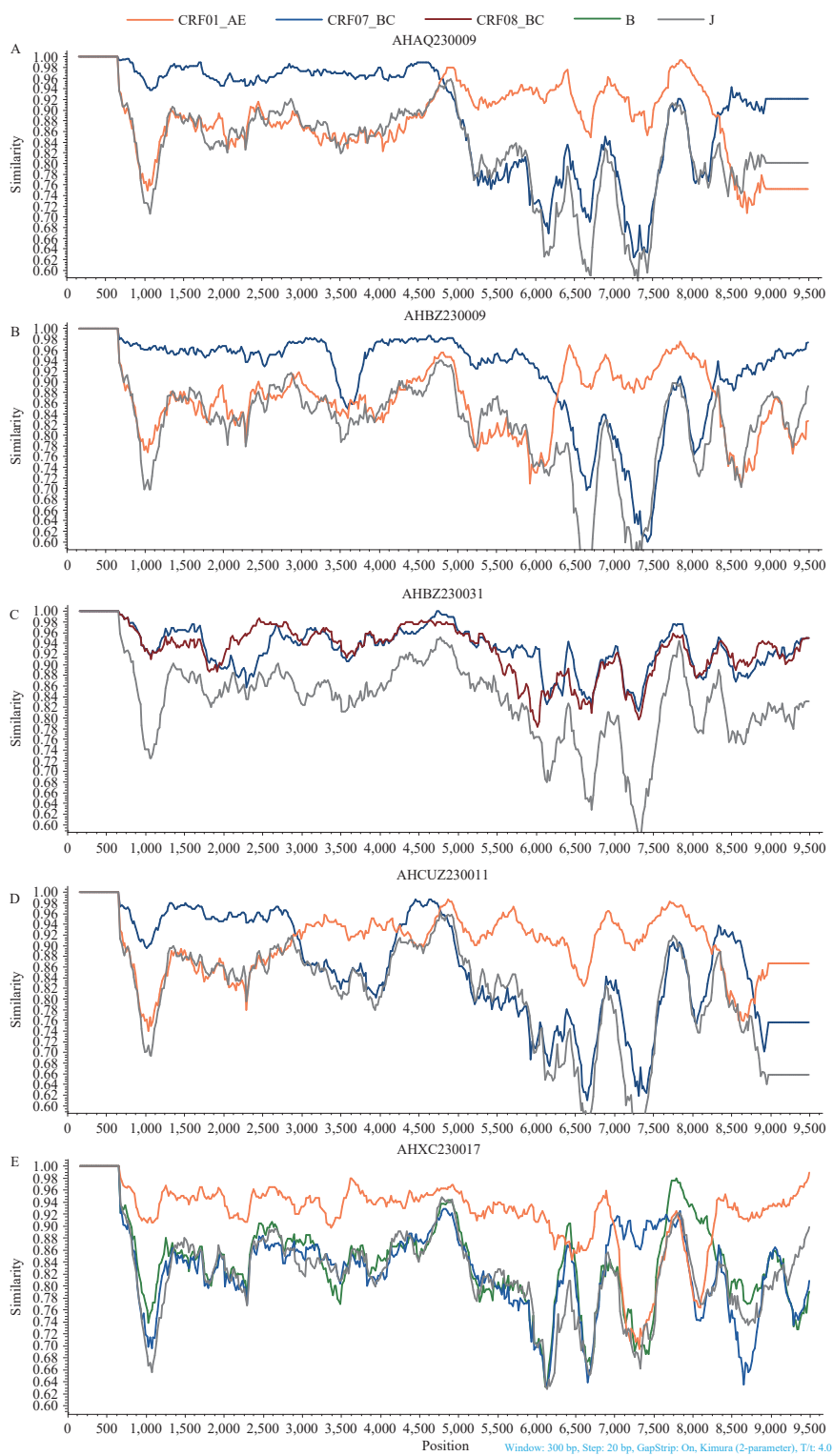
SUPPLEMENTARY TABLE S2. Subtype classification of genomic subregions from five near full-length genomes.

Fragments	AHAQ230009	AHBZ230009	AHBZ230031	AHCUZ230011	AHXC230017
I	CRF07_BC (634-4884)	CRF07_BC (634-6323)	CRF07_BC (620-1933)	CRF07_BC (608-2959)	CRF01_AE (621-4832)
II	CRF01_AE (4885-8404)	CRF01_AE (6324-8327)	CRF08_BC (1934-3001)	CRF01_AE (2960-4244)	B (4833-4899)
III	CRF07_BC (8405-8960)	CRF07_BC (8328-9552)	CRF07_BC (3002-3215)	CRF07_BC (4245-4885)	CRF01_AE (4900-6937)
IV			CRF08_BC (3216-4628)	CRF01_AE (4886-8327)	CRF07_BC (6938-7679)
V			CRF07_BC (4629-8518)	CRF07_BC (8328-8811)	B (7680-8240)
VI			CRF08_BC (8519-9098)	CRF01_AE (8812-9354)	CRF01_AE (8241-9552)
VII			CRF07_BC (9099-9470)		



SUPPLEMENTARY FIGURE S1. Phylogenetic subregion analysis of (A) AHAQ230009; (B) AHBZ230009; (C) AHBZ230031; (D) AHCUZ230011; and (E) AHXC230017.

Note: Subregion phylogenetic trees were constructed using IQ-TREE 2 software implementing the neighbor-joining method with 1,000 bootstrap replications. Scale bars represent 10% genetic distance. Figure E(II) is absent due to insufficient sequence length of the second fragment of AHXC230017 for phylogenetic reconstruction.



SUPPLEMENTARY FIGURE S2. Similarity analysis of five near full-length genomes. (A) AHAQ230009; (B) AHBZ230009; (C) AHBZ230031; (D) AHCUZ230011; and (E) AHXC230017.

Note: Analyses were performed using a 300-base window size and 20-base step size, aligned against HIV-1 subtype references. The CRF01_AE reference group included JX112859, EF036527, and AY358049; the CRF07_BC reference group included SC070050, HQ215552, and AF286230; the CRF08_BC reference group included KC914396 and AY008715; and the subtype B reference group included GU177863 and U71182. The subtype J reference was SE9280. The x-axis represents nucleotide positions, while the y-axis shows similarity values.