Vital Surveillances

National and Regional Molecular Epidemiology of HIV-1 — China, 2004–2023

Dong Wang[1]; Yi Feng[1]; Jingjing Hao[1]; Hongping Hu[1]; Fangyuan Li[1]; Jialu Li[1]; Yuhua Ruan[1]; Lingjie Liao[1]; Jing Hu[1]; Chang Song[1]; Yiming Shao[1]; Hui Xing[1,#]

ABSTRACT

Introduction: The genetic diversity of human immunodeficiency virus (HIV)-1 subtypes significantly influences the effectiveness of diagnostic tools, antiretroviral therapy (ART), and vaccine development. This study aimed to assess the regional and national prevalence of HIV-1 subtypes and recombinants in China between 2004 and 2023 using pol gene segment analysis.

Methods: We analyzed annual HIV/AIDS reports and pol gene segment sequences from all Chinese provinces between 2004 and 2023. The distribution of HIV-1 subtypes and recombinants across China and within its regions was estimated by multiplying the proportion of each subtype, circulating recombinant form (CRF), and unique recombinant form (URF) in each province by the corresponding number of reported HIV infections.

Results: Analysis of 94,476 pol gene segments from 31 provinces revealed that CRF01_AE strain accounted for 32.1% of HIV-1 infections during 2004–2023, while CRF07_BC lineage represented 39.1%. CRF08_BC strain contributed 9.2%, followed by subtype B (8.7%) and CRF55_01B (2.4%). Other CRFs collectively comprised 6.0% of infections, while URFs and other subtypes accounted for 1.3% and 1.1%, respectively.

Conclusions: The study revealed significant regional variations and temporal changes in the proportions of HIV-1 CRFs, subtypes, and URFs across China, emphasizing the importance of continued surveillance of strain distribution patterns.

Human immunodeficiency virus type 1 (HIV-1) remains a critical global health security challenge, affecting 39.9 million people as of 2023 (*1*). While expanded access to antiretroviral therapy (ART) has effectively reduced HIV-1 transmission rates, a definitive cure remains elusive, underscoring the critical importance of developing an effective HIV-1 vaccine for pandemic eradication (*2*).

The extensive genetic heterogeneity of HIV-1 presents significant challenges for diagnostics, ART, and vaccine development (*3*). According to the Los Alamos National Laboratory (LANL), HIV-1 group M strains comprise 10 distinct subtypes, along with circulating recombinant forms (CRFs) and unique recombinant forms (URFs) that arise through subtype recombination. Contemporary vaccine development strategies encompass mosaic vaccines, polyvalent approaches, and targeting evolutionarily conserved regions within the HIV genome. However, successful design, evaluation, and implementation of HIV-1 vaccines necessitate comprehensive and precise data regarding the prevalence of HIV-1 subtypes and recombinants to ensure vaccine immunogens match the circulating strains in target populations (*4*).

Continuous surveillance of strain distribution patterns is therefore essential. Global distribution patterns of HIV-1 subtypes and recombinants have been extensively documented by Hemelaar et al. (*5*) and Williams et al. (*6*), revealing CRF strains, particularly CRF01_AE, as predominant in East and Southeast Asia. While China similarly exhibits CRF strain predominance, it demonstrates distinct prevalence patterns characterized by the co-circulation of CRF01_AE, CRF07_BC, CRF08_BC, and CRF55_01B (*7*).

In this study, we analyzed pol gene segment sequences collected across China in conjunction with provincial HIV infection reports from 2004–2023 to comprehensively assess the regional and national distribution patterns of HIV-1 subtypes and recombinants.

METHODS

Our study analyzed a comprehensive dataset of HIV-1 pol region sequences from China comprising

two distinct subsets. The first subset consisted of sequences retrieved from the Los Alamos HIV Sequence Database, while the second originated from the Division of Research on Virology and Immunology at the National Center for AIDS/STD Control and Prevention, China CDC. Following rigorous quality control measures, including the removal of duplicate sequences and those lacking provincial sampling information, we established a refined dataset of 94,476 unique pol region sequences, ensuring one sequence per infected individual.

For sequence subtyping, we employed a dual-analysis approach. Initially, sequences were processed through the HIV subtyping tool of the National Microbiology Data Center's HIV database. Concurrently, we conducted phylogenetic analysis using FastTree. To resolve any discrepancies between these methodologies, we performed additional confirmatory analysis using IQtree2.exe, ensuring accurate subtype classification.

To estimate the national and regional distribution patterns of HIV-1 subtypes, we integrated our sequence analysis with provincial HIV infection surveillance data. The proportion of each subtype, CRF, and URF within each province was multiplied by the corresponding number of reported HIV infections to calculate comprehensive distribution estimates across China's regions and time periods.

RESULT

Characteristics of the Database

Our comprehensive analysis encompassed 94,476 HIV-1 pol gene segment sequences collected across China from 2004–2023.

Analysis of the sequence database revealed CRF01_AE as the predominant strain, comprising 38.6% (36,486/94,476) of all sequences, followed closely by CRF07_BC at 33.2% (31,388/94,476). Subtype B represented 9.6% (9,080/94,476) of sequences, while CRF08_BC and CRF55_01B accounted for 7.0% (6,575/94,476) and 4.3% (4,052/94,476), respectively. Notably, the aggregate proportion of all circulating recombinant forms (CRFs) constituted a substantial 88.5% (83,595/94,476) of the total sequences (Table 1, Supplementary Table S1, available at https://weekly.chinacdc.cn/).

Distribution of HIV-1 CRFs, Subtypes, and URFs in China

This extensive dataset was systematically stratified into four distinct chronological periods: 2004–2009, 2010–2014, 2015–2019, and 2020–2023, as detailed in Table 2. During 2004–2023, CRF01_AE and CRF07_BC emerged as the predominant strains, accounting for 32.1% and 39.1% of HIV-1 infections, respectively. CRF08_BC contributed significantly at 9.2%, followed by subtype B at 8.7% and CRF55_01B at 2.4%. Additional CRFs, including CRF59_01B, CRF79_0107, CRF85_BC, and 81 other distinct recombinant forms, collectively represented 6.0% of infections. URFs constituted 1.3% of cases, while other subtypes (including A1, C, D, and nine other subtypes or sub-subtypes) accounted for the remaining 1.1%.

Temporal analysis revealed significant shifts in HIV-1 subtype distribution (Table 2 and Figure 1).

TABLE 1. Distribution of HIV-1 subtypes, circulating recombinant forms (CRFs), and unique recombinant forms (URFs) in China.

Characteristics	HIV-1 CRFs						URFs	Total recombinants	HIV-1 subtypes		
	01_AE	07_BC	08_BC	55_01B	Other	Total CRFs			B	Other	Total subtypes
Weighted (%)	32.1	39.1	9.2	2.4	6.0	88.9	1.3	90.2	8.7	1.1	9.8
Unweighted (%)	38.6	33.2	7.0	4.3	5.4	88.5	1.1	89.6	9.6	0.8	10.4

Note: CRF denotes circulating recombinant form; URF denotes unique recombinant form. Other CRFs encompass CRF02_AG, CRF03_A6B, CRF06_cpx, CRF09_cpx, CRF10_CD, CRF11_cpx, CRF12_BF, CRF13_cpx, CRF15_01B, CRF16_A2D, CRF18_cpx, CRF19_cpx, CRF22_01A1, CRF25_cpx, CRF33_01B, CRF34_01B, CRF35_A1D, CRF43_02G, CRF45_cpx, CRF51_01B, CRF52_01B, CRF53_01B, CRF54_01B, CRF56_cpx, CRF57_BC, CRF58_01B, CRF59_01B, CRF61_BC, CRF62_BC, CRF64_BC, CRF65_cpx, CRF67_01B, CRF68_01B, CRF69_01B, CRF76_01B, CRF77_cpx, CRF79_0107, CRF80_0107, CRF84_A1D, CRF85_BC, CRF87_cpx, CRF88_BC, CRF96_cpx, CRF97_01B, CRF100_01C, CRF101_01B, CRF102_0107, CRF103_01B, CRF104_0107, CRF105_0108, CRF106_cpx, CRF107_01B, CRF109_0107, CRF110_BC, CRF111_01C, CRF112_01B, CRF113_0107, CRF114_0155, CRF115_01C, CRF118_BC, CRF119_0107, CRF120_0107, CRF121_0107, CRF123_0107, CRF125_0107, CRF126_0755, CRF128_07B, CRF134_0107, CRF137_0107, CRF140_0107, CRF145_0107, CRF149_01B, CRF151_0107, CRF154_0755, CRF155_0755. Other subtypes include subtypes A1, A3, A6, C, D, F1, F2, G, and H. The unweighted row represents the actual proportions of subtypes, CRFs, and URFs in the sequence database. The weighted row presents estimated proportions based on actual sequence distributions and provincial HIV infection reports.Total CRFs is the sum of CRF01_AE, CRF07_BC, CRF08_BC, CRF55_01B, and other CRFs. Total recombinants is the sum of total CRFs and URFs. Total subtypes is the sum of subtype B and other subtypes.

TABLE 2. Spatiotemporal distribution of HIV-1 subtypes, CRFs, and URFs across four time periods.

Region/Years	HIV-1 CRFs						URFs	Total recombinants	HIV-1 subtypes		
	01_AE	07_BC	08_BC	55_01B	Other	Total CRFs			B	Other	Total subtypes
China											
2004–2009 (%)	31.0	24.8	10.0	0.6	5.0	71.4	1.7	73.1	24.7	2.2	26.9
2010–2014 (%)	41.1	35.9	6.0	2.1	5.6	90.7	0.9	91.6	7.2	1.3	8.4
2015–2019 (%)	31.4	46.1	8.4	3.0	4.4	93.4	0.9	94.3	5.0	0.7	5.7
2020–2023 (%)	27.6	42.2	11.9	3.2	9.0	94.0	1.8	95.9	3.5	0.6	4.1
Central											
2004–2009 (%)	12.8	1.8	0.3	0.1	0.4	15.4	0.1	15.5	84.3	0.2	84.5
2010–2014 (%)	44.1	20.8	1.9	4.4	3.1	74.4	0.2	74.6	25.1	0.3	25.4
2015–2019 (%)	37.1	30.0	4.4	5.7	3.2	80.3	1.1	81.4	18.1	0.5	18.6
2020–2023 (%)	30.1	37.7	5.9	3.7	8.3	85.8	1.8	87.5	11.3	1.2	12.5
East											
2004–2009 (%)	38.5	16.5	3.4	0.0	2.6	61.0	0.2	61.2	36.8	2.0	38.8
2010–2014 (%)	56.0	25.7	1.7	1.4	5.6	90.2	0.2	90.4	7.1	2.4	9.6
2015–2019 (%)	40.1	35.0	5.1	3.6	6.4	90.2	1.1	91.3	7.9	0.8	8.7
2020–2023 (%)	30.4	40.1	6.3	2.8	11.8	91.5	2.8	94.3	4.4	1.3	5.7
North											
2004–2009 (%)	35.1	9.1	1.6	0.6	0.8	47.2	0.1	47.3	52.3	0.4	52.7
2010–2014 (%)	52.1	21.7	1.1	1.1	4.4	80.3	0.3	80.6	18.7	0.7	19.4
2015–2019 (%)	52.3	29.7	0.9	2.4	6.2	91.4	0.8	92.2	7.5	0.3	7.8
2020–2023 (%)	37.3	38.7	1.0	2.2	13.1	92.3	2.7	95.0	4.9	0.0	5.0
Northeast											
2004–2009 (%)	36.4	2.1	5.3	0.3	8.2	52.2	0.0	52.2	47.2	0.6	47.8
2010–2014 (%)	71.6	9.3	0.0	0.8	5.9	87.6	0.9	88.5	9.4	2.1	11.5
2015–2019 (%)	71.5	16.9	0.5	0.8	4.1	93.8	0.3	94.1	5.5	0.4	5.9
2020–2023 (%)	48.4	29.7	0.8	1.5	10.3	90.8	2.7	93.4	5.6	1.0	6.6
Northwest											
2004–2009 (%)	0.9	93.6	0.1	0.0	0.0	94.6	0.0	94.6	5.3	0.0	5.4
2010–2014 (%)	16.1	77.9	0.1	0.4	1.0	95.5	0.3	95.8	4.2	0.0	4.2
2015–2019 (%)	15.0	74.5	0.7	2.1	3.4	95.6	0.6	96.3	3.4	0.3	3.7
2020–2023 (%)	10.8	67.6	1.9	1.6	11.7	93.6	3.7	97.3	2.0	0.7	2.7
South											
2004–2009 (%)	69.2	13.5	6.3	2.3	2.6	93.9	0.3	94.2	4.7	1.1	5.8
2010–2014 (%)	62.3	20.6	7.7	4.6	1.9	97.1	0.2	97.3	2.4	0.2	2.7
2015–2019 (%)	54.6	24.2	8.2	7.7	2.8	97.5	0.4	97.9	1.9	0.2	2.1
2020–2023 (%)	39.1	26.3	13.7	9.9	6.8	95.8	2.2	98.1	1.7	0.3	1.9
Southwest											
2004–2009 (%)	21.0	29.0	22.5	0.0	10.9	83.4	4.5	87.9	7.4	4.7	12.1
2010–2014 (%)	21.9	50.8	10.5	0.8	9.6	93.7	1.9	95.6	2.5	1.9	4.4
2015–2019 (%)	13.7	63.4	14.1	0.9	4.6	96.7	1.1	97.9	1.0	1.1	2.1
2020–2023 (%)	20.2	48.9	19.5	0.9	7.9	97.4	0.8	98.2	1.3	0.4	1.8

Note: The Northeast region comprises HLJ, JL, and LN; North includes NM, BJ, TJ, HE, and SX; East encompasses SH, JS, ZJ, AH, FJ, JX, and SD; South consists of GD, GX, and HI; Central includes HA, HB, and HN; Northwest comprises SN, GS, QH, NX, and XJ; Southwest includes CQ, SC, GZ, YN, and XZ. Data are presented as proportions of total reported HIV infections within each region and time period.

Abbreviation: HIJ=Heilongjiang; JL=Jilin; LN=Liaoning; NM=Neimenggu; BJ=Beijing; TJ=Tianjin; HE=Hebei; SX=Shanxi; SH=Shanghai; JS=Jiangsu; ZJ=Zhejiang; AH=Anhui; FJ=Fujian; JX=Jiangxi; SD=Shandong; GD=Guangdong; GX=Guangxi; HI=Hainan; HA=Henan; HB=hubei; HE=Hunan; SN=Shannxi; GS=Gansu; QH=Qinghai; NX=Ningxia; XJ=Xinjiang; CQ=Chongqing; SC=Sichuan; GZ=Guizhou; YN=Yunnan; XZ=Xizang.
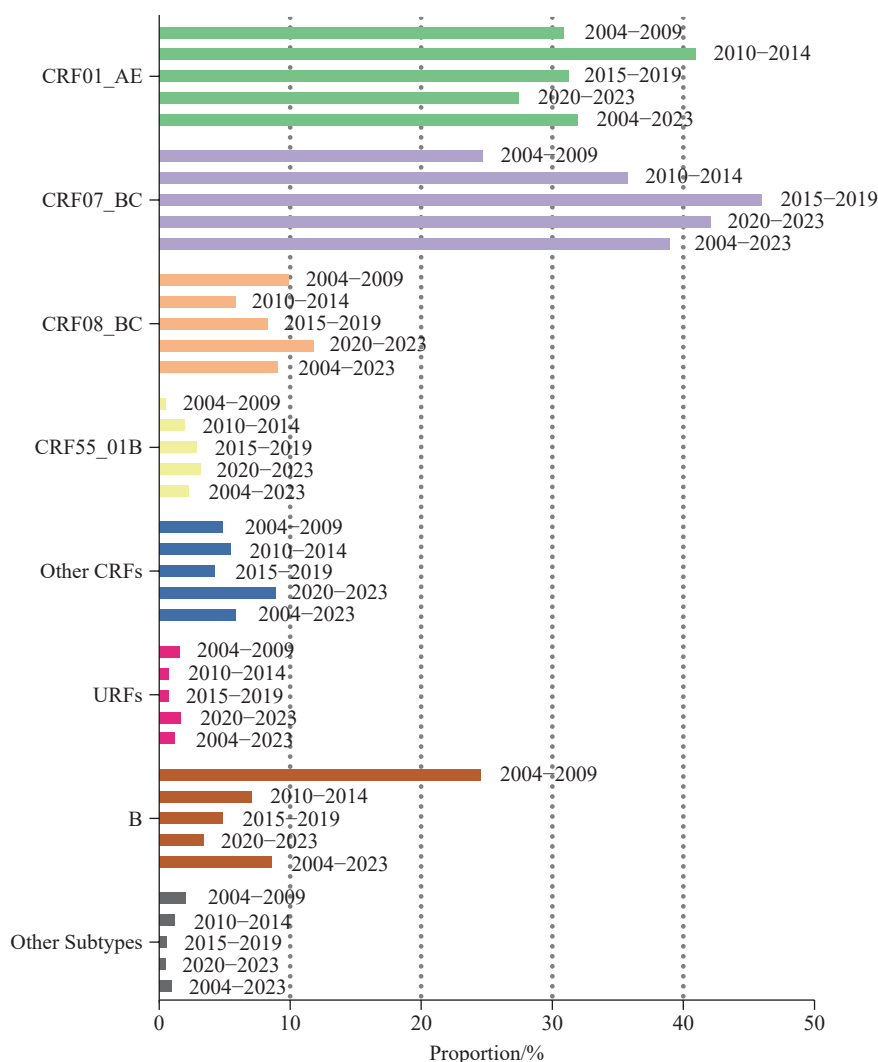
FIGURE 1. Temporal trends in the distribution of HIV-1 genetic variants in China, 2004–2023.
abbreviation: HIV-1=human immunodeficiency virus type 1.

CRF07_BC demonstrated a marked increase from 24.8% to 46.1% during 2004–2019, followed by a decline in 2020–2023. Conversely, CRF01_AE reached its peak prevalence of 41.1% during 2010-2014, subsequently showing a gradual decrease through 2015–2023. Subtype B exhibited a consistent downward trend, reaching its lowest prevalence of 3.5% in 2020–2023.

The epidemiological pattern of CRF08_BC showed initial decline followed by steady increase from 2015–2023. CRF55_01B demonstrated consistent growth throughout the study period, increasing from 0.6% to 3.2%. Other CRFs showed fluctuating patterns, reaching their highest prevalence of 9.0% during 2020–2023.

URF prevalence exhibited a complex pattern, initially decreasing through the first two time periods,

stabilizing during 2010–2019, and subsequently increasing in 2020–2023. In contrast, other subtypes showed a consistent declining trend throughout the entire study period (2004–2023).

Regional Distribution of HIV-1 CRFs, Subtypes, and URFs

The proportions of HIV-1 CRFs, subtypes, and URFs exhibited distinct regional variations across China's seven geographical regions and demonstrated significant temporal changes throughout the study period (Figure 2 and Table 2).

During 2004–2009, distinct regional patterns emerged: the northeastern, northern, and eastern regions were predominantly affected by CRF01_AE and subtype B strains, collectively accounting for over 75% of infections. The southern region showed
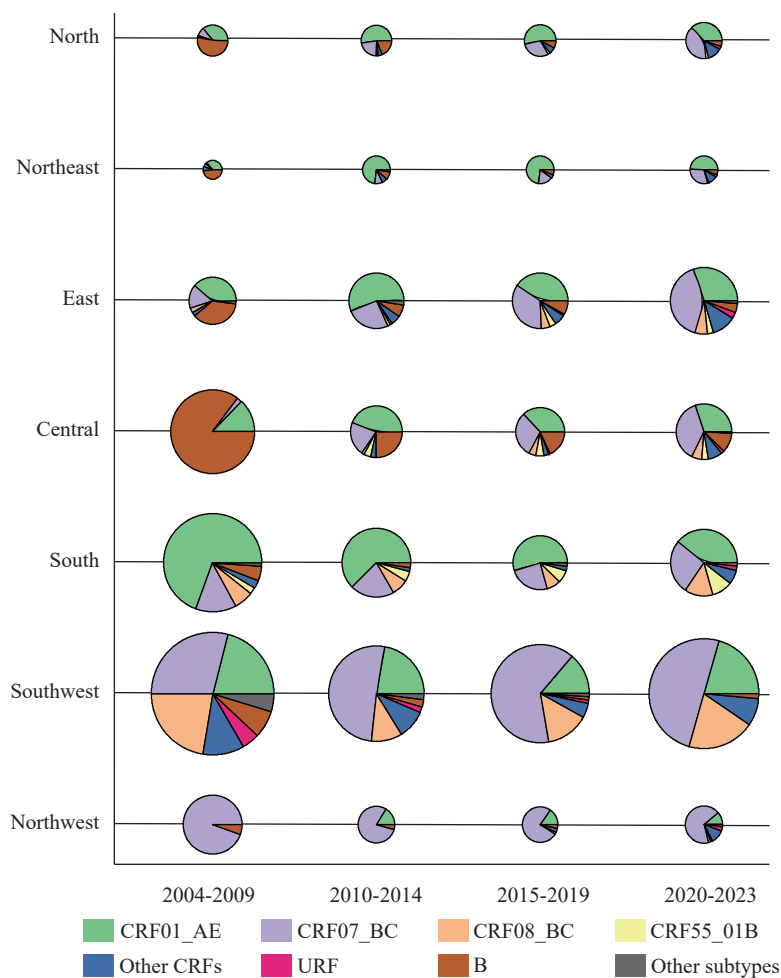
FIGURE 2. Regional Distribution and temporal evolution of HIV-1 genetic variants across China's seven regions, 2004–2009, 2010–2014, 2015–2019, and 2020–2023.

CRF01_AE dominance at 69.2%, while the central region exhibited a marked prevalence of subtype B (84.3%). The northwestern region was characterized by CRF07_BC predominance at 93.6%. The southwestern region displayed a more heterogeneous distribution, with CRF07_BC (29.0%), CRF08_BC (22.5%), and CRF01_AE (21.0%) collectively representing over 70% of infections.

A significant epidemiological shift occurred during 2020–2023, with CRF07_BC and CRF01_AE emerging as the predominant strains across all regions.

The study period revealed several notable trends: total recombinant forms showed consistent increases across six regions, with the Northeast being the sole exception. Similarly, CRF07_BC demonstrated sustained growth in all regions except the Northwest and Southwest. The 2020–2023 period was marked by regional distinctions: the Northeast showed the highest CRF01_AE prevalence (48.4%), the North recorded the highest proportion of other CRFs (13.1%), and the

South maintained the highest CRF55_01B prevalence throughout the entire study period.

Central China maintained a higher proportion of subtype B infections compared to other regions throughout 2004–2023, despite its declining trend. The Northwest's epidemiological profile was dominated by CRF07_BC, which decreased from approximately 95% (2004–2009) to 67.6% (2020–2023). The Southwest consistently maintained the highest proportion of CRF08_BC infections across all four time periods.

CONCLUSIONS

This study represents the most comprehensive analysis to date of HIV-1 genetic diversity in China, encompassing 94,476 samples over a 20-year period (2004–2023). Our findings demonstrate that CRF07_BC and CRF01_AE are the predominant viral

lineages, with significant contributions from CRF08_BC, CRF55_01B, and subtype B to the overall genetic landscape. Notably, circulating recombinant forms (CRFs) account for 88.6% of all HIV-1 infections, underscoring their central role in China's HIV epidemic.

The epidemiological landscape of HIV-1 in China has undergone substantial transformation over the past decade, characterized by distinct temporal dynamics in subtype distribution. While the prevalence of CRF01_AE, CRF07_BC, and subtype B has declined, other subtypes have maintained relative stability. Conversely, CRF08_BC, CRF55_01B, URFs, and other CRFs have shown increasing prevalence. These evolutionary patterns reflect the complex interplay of multiple factors and the unique trajectory of HIV-1 transmission in China.

The early epidemic phase was marked by widespread outbreaks of subtype B strains within the BLD (8). Subsequent implementation of stringent blood product controls effectively reduced subtype B prevalence across central China and nationwide. Despite reintroduction into the HET population, subtype B strains have continued to show declining trends (9).

The epidemiological landscape has been further shaped by distinct viral lineages: CRF01_AE strains, introduced from the Golden Triangle region, have established varying prevalence patterns through multiple transmission events across different regions (10–11). CRF07_BC, originating in YN before spreading to XJ and SC, achieved particular prominence in northwestern and southwestern regions. Initially associated with IDU and HET transmission routes, a distinct CRF07_BC lineage emerged around 2005, predominantly spreading among MSM and leading to substantial increases in prevalence across central and eastern provinces (12).

CRF08_BC, which originated among injection drug users in Yunnan Province and was subsequently identified in Guangxi, has become a predominant strain in Southwest China (13). While its prevalence initially decreased due to declining injection drug use, the strain's successful transition into heterosexual transmission networks has led to a resurgence. After 2009, when sexual transmission accounted for over 90% of HIV infections in China, the epidemiological dynamics of CRF08_BC were further altered by these changing transmission patterns.

The CRF55_01B strain, which emerged in Guangdong Province, has achieved nationwide distribution primarily through China's railway infrastructure, particularly along the Beijing-Guangzhou and Beijing-Kowloon corridors. Despite its broad dissemination, this strain maintains notably higher prevalence rates in South China (7,14).

These temporal and geographical factors have collectively shaped China's complex HIV-1 epidemiological landscape, emphasizing the necessity for regionally tailored prevention and control strategies that account for these variations.

The distinct geographical distributions of HIV-1 lineages in China are significantly influenced by their biological characteristics. Research has revealed substantial differences between CRF01_AE and CRF07_BC strains, particularly in disease progression rates and X4-tropic strain prevalence. The heterogeneous biological properties of CRF01_AE lineages contribute to variable disease progression rates, subsequently affecting their transmission dynamics and distribution patterns (15). Studies indicate that CRF07_BC's emergence as a predominant strain in China may be attributed to its slower replication rate, which correlates with delayed disease progression (16). Additionally, CRF55_01B's rapid dissemination might be explained by its natural polymorphism at position V179 (17), associated with NVP and EFV resistance, resulting in higher viral loads and more rapid CD4 T-cell depletion during early infection (18).

Our comprehensive analysis of 94,476 sequences spanning 2004–2023 revealed significant geographical and sociocultural influences on HIV-1 epidemiology in China. A consistent increase in recombinant forms was observed across all regions except central China, with recombinants accounting for over 90% of infections in most regions during 2020–2023. The emergence of novel CRFs in recent years indicates ongoing recombination events, emphasizing the critical need for enhanced surveillance measures.

Several limitations warrant consideration in our study. Although some provinces initially provided limited sequence data, their relatively low HIV infection rates suggest minimal impact on the overall regional and national proportions. Additionally, our reliance on pol genome segment analysis may have resulted in an underestimation of URF prevalence.

The genetic diversity of HIV-1 subtypes significantly impacts diagnostic accuracy, ART efficacy, and vaccine development strategies. The selection of appropriate vaccine immunogen sequences presents a considerable challenge, necessitating precise alignment with circulating strains. Known variations in diagnostic efficiencies across subtypes and recombinants, coupled

with specific mutations like V179 in CRF55_01B, underscore the importance of continuous surveillance for drug-resistant mutations across all viral variants. This ongoing monitoring is crucial for optimizing both therapeutic approaches and diagnostic methodologies.

In conclusion, our study provides an essential comprehensive analysis of HIV-1 subtype and recombinant distribution patterns in China. These findings have direct implications for HIV vaccine development, diagnostic reagent selection, and ART regimen optimization. The dynamic nature of HIV-1 genetic diversity emphasizes the necessity for continued surveillance of subtype and recombinant proportions to inform and enhance public health strategies.

Corresponding author: Hui Xing, xingh@chinaaids.cn.

---

¹ National Center for AIDS/STD Control and Prevention (NCAIDS), Chinese Center for Disease Control and Prevention (China CDC), State Key Laboratory of Infectious Disease Prevention and Control (SKLID), Beijing, China.

REFERENCES

1. UNAIDS. Global AIDS update 2024，https://www.unaids.org, [2024].
2. Fauci AS. An HIV vaccine is essential for ending the HIV/AIDS pandemic. JAMA 2017;318(16):1535 – 6. https://doi.org/10.1001/jama.2017.13505.
3. Gregson J, Tang M, Ndembi N, Hamers RL, Rhee SY, Marconi VC, et al. Global epidemiology of drug resistance after failure of WHO recommended first-line regimens for adult HIV-1 infection: a multicentre retrospective cohort study. Lancet Infect Dis 2016;16(5): 565 – 75. https://doi.org/10.1016/S1473-3099(15)00536-8.
4. Gaschen B, Taylor J, Yusim K, Foley B, Gao F, Lang D, et al. Diversity considerations in HIV-1 vaccine selection. Science 2002;296(5577): 2354 – 60. https://doi.org/10.1126/science.1070441.
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. Williams A, Menon S, Crowe M, Agarwal N, Biccler J, Bbosa N, et al. Geographic and population distributions of human immunodeficiency virus (HIV)–1 and HIV-2 circulating subtypes: a systematic literature review and meta-analysis (2010–2021). J Infect Dis 2023;228(11):1583 – 91. https://doi.org/10.1093/infdis/jiad327.
7. Hao JJ, Zheng S, Gan MZ, Dong AB, Kang RH, Li MM, et al. Changing proportions of HIV-1 subtypes and transmitted drug resistance among newly diagnosed HIV/AIDS individuals - China, 2015 and 2018. China CDC Wkly 2021;3(53):1133 – 8. https://doi.org/10.46234/ccdcw2021.251.
8. Li Z, He X, Wang Z, Xing H, Li F, Yang Y, et al. Tracing the origin and history of HIV-1 subtype B′ epidemic by near full-length genome analyses. AIDS 2012;26(7):877 – 84. https://doi.org/10.1097/QAD.0b013e328351430d.
9. Li L, Sun GQ, Zhong P, Han JW, Li TY, Jia DJ, et al. HIV‑1 Thai B strain has spread out of former plasma donors into general population through sexual contact in Henan, China. J Med Virol 2016;88(4):614 – 21. https://doi.org/10.1002/jmv.24383.
10. Feng Y, He X, Hsi JH, Li F, Li XG, 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. https://doi.org/10.1097/QAD.0b013e328360db2d.
11. Wang XL, He X, Zhong P, Liu YJ, Gui T, Jia DJ, et al. Phylodynamics of major CRF01_AE epidemic clusters circulating in mainland of China. Sci Rep 2017;7(1):6330. https://doi.org/10.1038/s41598-017-06573-6.
12. Gan MZ, Zheng S, Hao JJ, Ruan YH, Liao LJ, Shao YM, et al. Spatiotemporal patterns of CRF07_BC in China: a population-based study of the HIV strain with the highest infection rates. Front Immunol 2022;13:824178. https://doi.org/10.3389/fimmu.2022.824178.
13. 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.
14. Gan MZ, Zheng S, Hao JJ, Ruan YH, Liao LJ, Shao YM, et al. The prevalence of CRF55_01B among HIV-1 strain and its connection with traffic development in China. Emerg Microbes Infect 2021;10(1):256 – 65. https://doi.org/10.1080/22221751.2021.1884004.
15. Li K, Chen HH, Li JJ, Feng Y, Liang SJ, Rashid A, et al. Distinct genetic clusters in HIV-1 CRF01_AE-infected patients induced variable degrees of CD4⁺ T-cell loss. mBio 2024;15(3):e0334923. https://doi.org/10.1128/mbio.03349-23.
16. Ma LY, Guo YF, Yuan L, Huang Y, Sun JP, Qu SL, et al. Phenotypic and genotypic characterization of human immunodeficiency virus type 1 CRF07_BC strains circulating in the Xinjiang province of China. Retrovirology 2009;6:45. https://doi.org/10.1186/1742-4690-6-45.
17. Liu YJ, Li HP, Wang XL, Han JW, Jia L, Li TY, et al. Natural presence of V179E and rising prevalence of E138G in HIV-1 reverse transcriptase in CRF55_01B viruses. Infect Genet Evol 2020;77: 104098. https://doi.org/10.1016/j.meegid.2019.104098.
18. Wei L, Li H, Lv X, Zheng CL, Li GL, Yang ZR, et al. Impact of HIV-1 CRF55_01B infection on the evolution of CD4 count and plasma HIV RNA load in men who have sex with men prior to antiretroviral therapy. Retrovirology 2021;18(1):22. https://doi.org/10.1186/s12977-021-00567-z.

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY TABLE S1. Database Composition: Distribution of HIV-1 genetic variants across 2004–2009, 2010–2014, 2015–2019, and 2020–2023.

Region/Years	HIV-1 CRFs					URF	Total CRFs	Total recombinants	HIV-1 Subtypes		Total Subtypes	N
	01_AE	07_BC	08_BC	55_01B	Other				B	Other		
China												
2004–2009	29.6%	22.3%	5.2%	1.3%	3.5%	0.9%	61.9%	62.8%	35.4%	1.8%	37.2%	9,488
2010–2014	44.7%	30.7%	4.5%	4.5%	4.4%	0.9%	88.8%	89.7%	9.2%	1.1%	10.3%	22,921
2015–2019	41.4%	34.1%	6.7%	5.5%	4.7%	0.9%	92.4%	93.2%	6.3%	0.5%	6.8%	43,480
2020–2023	29.2%	39.8%	11.4%	2.9%	9.2%	2.0%	92.5%	94.5%	4.8%	0.7%	5.5%	18,587
Central												
2004–2009	9.8%	1.5%	0.2%	0.0%	0.4%	0.1%	12.0%	12.0%	87.8%	0.2%	88.0%	2,225
2010–2014	51.9%	21.7%	2.4%	4.6%	2.8%	0.3%	83.4%	83.7%	16.2%	0.1%	16.3%	711
2015–2019	36.0%	29.0%	3.8%	5.6%	3.1%	1.0%	77.5%	78.4%	21.1%	0.4%	21.6%	2,662
2020–2023	29.0%	37.2%	3.9%	3.9%	7.1%	1.6%	81.0%	82.6%	16.6%	0.7%	17.4%	2,269
East												
2004–2009	35.3%	12.3%	5.0%	0.0%	3.0%	0.6%	55.6%	56.2%	41.3%	2.5%	43.8%	1,509
2010–2014	58.7%	25.6%	0.9%	1.6%	4.4%	0.2%	91.2%	91.4%	7.3%	1.3%	8.6%	2,475
2015–2019	41.5%	29.7%	4.1%	3.8%	6.8%	1.2%	85.8%	87.1%	12.4%	0.5%	12.9%	8,510
2020–2023	31.2%	40.0%	6.5%	3.4%	10.5%	2.6%	91.7%	94.3%	4.7%	1.0%	5.7%	5,298
North												
2004–2009	37.5%	15.7%	3.2%	1.2%	1.7%	0.2%	59.4%	59.6%	39.6%	0.8%	40.4%	997
2010–2014	50.7%	24.3%	1.1%	1.1%	4.0%	0.3%	81.2%	81.5%	17.9%	0.6%	18.5%	6,637
2015–2019	50.1%	30.3%	0.9%	2.5%	6.4%	0.8%	90.2%	90.9%	8.8%	0.3%	9.1%	5,536
2020–2023	37.7%	38.0%	1.3%	2.5%	13.8%	2.1%	93.3%	95.4%	4.5%	0.1%	4.6%	1,734
Northeast												
2004–2009	66.9%	4.3%	1.8%	0.6%	8.0%	0.0%	81.6%	81.6%	17.2%	1.2%	18.4%	163
2010–2014	77.7%	6.7%	0.0%	0.3%	6.1%	1.1%	90.8%	91.9%	7.0%	1.1%	8.1%	358
2015–2019	72.2%	16.6%	0.4%	0.8%	4.0%	0.2%	94.0%	94.2%	5.5%	0.3%	5.8%	1,264
2020–2023	51.3%	29.6%	0.6%	1.6%	8.8%	1.9%	92.0%	93.9%	5.3%	0.8%	6.1%	622
Northwest												
2004–2009	4.0%	85.1%	0.3%	0.0%	0.0%	0.0%	89.4%	89.4%	10.4%	0.2%	10.6%	597
2010–2014	21.4%	71.3%	0.2%	0.3%	1.1%	0.3%	94.2%	94.5%	5.5%	0.0%	5.5%	637
2015–2019	18.4%	69.1%	0.8%	2.4%	4.1%	0.8%	94.8%	95.6%	4.1%	0.3%	4.4%	1,109
2020–2023	10.5%	70.9%	1.5%	1.6%	9.9%	3.0%	94.5%	97.5%	2.0%	0.5%	2.5%	910
South												
2004–2009	53.5%	21.9%	5.6%	4.7%	3.7%	0.4%	89.5%	89.8%	8.1%	2.1%	10.2%	2,220
2010–2014	43.5%	34.2%	3.8%	9.5%	2.9%	0.4%	93.9%	94.3%	5.1%	0.6%	5.7%	8,893
2015–2019	49.5%	28.3%	6.9%	9.3%	3.1%	0.4%	97.1%	97.5%	2.3%	0.2%	2.5%	17,808
2020–2023	39.3%	33.3%	8.8%	6.0%	7.8%	2.0%	95.1%	97.1%	2.6%	0.3%	2.9%	2,490
Southwest												
2004–2009	20.1%	41.6%	14.3%	0.1%	9.4%	3.9%	85.5%	89.4%	6.5%	4.1%	10.6%	1,777
2010–2014	24.0%	34.9%	18.3%	1.0%	10.6%	4.0%	88.8%	92.7%	3.6%	3.6%	7.3%	3,210
2015–2019	12.7%	58.2%	17.7%	1.0%	5.7%	1.6%	95.3%	96.9%	1.3%	1.8%	3.1%	6,591
2020–2023	20.1%	40.3%	27.2%	0.9%	8.0%	1.2%	96.6%	97.8%	1.3%	0.9%	2.2%	5,264